HELSINGIN YLIOPISTO

Modal Epistemology

A study of the conditions of knowledge

Jaakko Hirvelä

Pro gradu – tutkielma

Teoreettinen filosofia

Filosofian historian kulttuurin

ja taiteiden tutkimuksen laitos

Helsingin yliopisto

Helmikuu 2014

Table of Contents

1. Introduction

What is knowledge? Do we know anything at all? If we do, then how have we obtained our knowledge? Why do we hold knowledge in such a high esteem? Questions such as these troubled the ancient philosophers and continue to trouble us to this very day. Plato thought that knowledge is, basically, justified true belief (JTB). This traditional analysis of knowledge was widely held until refuted in 1963 by Edmund Gettier in his groundbreaking article "Is Justified True Belief Knowledge?" Analytic epistemology was in turmoil. What the Gettier cases showed was that a subject, S, could have a justified true belief, *p*, even though he did not know that *p*. The reason why the subjects of Gettier cases seemed to lack knowledge was because their beliefs seemed to be only luckily true. Good epistemic luck undermined the subject's knowledge but not her being justified. Being justified in a proposition didn't exclude the chance that one held one's belief simply as a matter of luck. Knowledge and justification seem to come apart when one's justified belief is true in virtue of good luck.

Nowadays it is generally accepted that luck - at least luck of a certain sort - is incompatible with knowledge. Let us call this intuition the *anti-luck intuition*. The main goal of this thesis is to evaluate whether by eliminating epistemic luck, and thus by satisfying the anti-luck intuition, we can arrive at a plausible theory of knowledge. The notion of 'justification' has proved to be insufficient as a necessary and sufficient condition for knowledge. Here we shall abandon it altogether and start afresh. Could knowledge be defined simply as a non-luckily true belief? That is the question that we will pursue.

The structure of this study is four folded. I will begin by clarifying the concept of epistemic luck, which is at the heart of this project. I will proceed to evaluate two widely discussed modal conditions that aim to eliminate epistemic luck. First in line is the 'sensitivity' condition, originally put forth by Robert Nozick. After discussing the merits of 'sensitivity' at length, I will turn to examine another modal condition, 'safety'. Critical examination of 'safety' shows that it too is susceptible to a number of counter examples, just as 'sensitivity' was. In the fourth and final section I argue that the explanatory power of 'safety' has not been properly understood. I present and defend a novel modal condition that is developed in this study from the ashes of the 'safety'

condition. The new condition, ‘global safety’ fares much better than its predecessors regarding a number of epistemological problems. ‘Global safety’ does not need to be conjoined with other conditions, and can therefore stand on its own feet. Thus I end up defending a pure version of the ‘safety’ condition as giving the necessary and sufficient conditions of knowledge. This position has been quite neglected in the contemporary epistemology.

2. Theory of knowledge

The goal of this thesis is to formulate a normative theory of knowledge that expresses the necessary conditions for knowledge and hopefully the sufficient conditions also. But in order to be able to evaluate whether the project succeeds at the end, we have to know what kind of theory would be satisfactory in the first place. What questions will a satisfactory theory of knowledge be able to answer and what kind of phenomena will it take into account?

These questions might have an obvious answer: ‘The theory should explain what knowledge is and it should handle epistemic phenomena’. Though this is true, the answer is hardly informative. What we really want from a theory of knowledge is that it fits well together with our pre-theoretical conception of knowledge. A successful theory will not be all too revisionistic. It should confirm most of our knowledge attributions as true and provide an adequate explanation for our erroneous attributions. Since we attribute knowledge to each other daily, a successful theory should explain how knowledge is possible in the first place. Ideally it would also account for the intuitive judgment that we don’t know the denials of skeptical hypotheses. The theory should also give an explanation why our conception of knowledge is what it is. Is there some reason why the verb ‘know’ has the meaning that it has in our society? Finally the theory should explain why knowledge is valuable, or if it is not, then it should provide an explanation why it seems to be of value. So the theory should be:

1. Accurate
2. Confirm most of our knowledge attributions as true
3. Explain how knowledge is possible
4. Explain why knowledge has this particular structure
5. Explain why knowledge is valuable

Frankly it all reduces to point 1. If the theory is accurate then it will provide the explanations required by points 2, 3, 4 and 5. What we have to find out is what will make a theory accurate. Furthermore there are two additional criteria that any adequate definition of knowledge must pass: the definition should be informative and non-circular. If I am able to provide a definition of knowledge that is accurate, informative and non-circular, then the project that has been undertaken here can be considered as a success. But even if we fail in this, we might have shed some light on the concept of knowledge, which would also be satisfactory result. Now to the task at hand!

3. Internalism and externalism

As I said in the introduction, the theories that I am going to discuss and advocate, do not attempt to explicate the notion of justification. The main reason for this is that the conception of justified true belief caused us all the trouble we've had since Gettier published his counter examples to the traditional analysis of knowledge. The reason why Gettier examples were possible was because justification was not appropriately tied to the truth of the belief.

Traditionally justification is understood as an internal phenomenon. According to internalism, justification of a belief is dependent solely on the internal states of the believer. But if internalism is true, then justification is not tied to truth in a robust enough way. Internalistic justification does not tie belief to truth in a reliable way. Externalists about justification tried to mend this problem. They claimed that in order to be justified one has to satisfy at least some external conditions. Reliabilists claimed that beliefs which are reliably formed are justified. According to reliabilists justified beliefs are much more likely to be true than unjustified ones. The problem with externalism about justification is that they could no longer accommodate some highly intuitive claims about justification. Internalism could easily explain why two subjects who are in identical mental states are equally justified about the propositions which they believed. Externalists could not accommodate this result because two subjects could be mentally alike and yet differ in the way they were related to the external world. Intuitively you and your envatted twin are equally justified in believing any proposition you believe in. Crucially however, only you are related to the external world in a reliable way, so that

most of your justified beliefs will be true, whereas most of your envatted twin's beliefs will be false.

So internalism about justification makes an account of knowledge susceptible to the Gettier cases while externalism about justification conflicts with the very notion of justification. The easiest route out of this problem is to abandon the notion of justification altogether. Because the reason why the JTB account of knowledge failed was that justification wasn't appropriately tied to the truth of the belief, it is clear that the condition that aims to replace justification should express a relation between the subject's belief and the truth of that belief. Does that mean that internalism is out of the window? Basically yes. Knowledge cannot be dependent on the truth of the belief and internal factors alone. The subject has to be appropriately related to the truth of the belief in order for her to possess knowledge. The theories that I'm going to look at will thus be externalist in nature. Of course even an externalist will want to accommodate the internalist's intuitions, or at least be able to explain why they are false. I will highlight places where one could opt for an internalistic solution or add an internal condition to accommodate internalistic intuitions. Furthermore, it is compatible with externalism that in addition to the external criteria there are some internal conditions that a subject must satisfy in order to have knowledge. This thesis is devoted to finding the external criteria.

4. Methods

In this thesis I will be using two, quite general methods of inquiry. Following the trend in contemporary epistemology I will be relying heavily on the counter example method. The conditions of knowledge that I study are modal in nature, and to clarify the modalities involved I will use the conception of possible worlds. Let me start with describing and motivating the counter example method.

4.1. Counter example method

The theory that we aim to put forward has to be in synch with our everyday knowledge attributions. It has to accord with our intuitions about knowledge. I aim to accomplish this feat by using the counter example method.

The post-Gettier epistemology is crowded by counter examples that undermine even the most elegant theories of knowledge. The counter example method is quite simple: Devise a case where a subject believes some true proposition *p* and reflect on your intuitions in order to determine whether the subject knows that *p*. If according to your intuitions the subject lacks/has knowledge in the case, but a theory of knowledge would dictate otherwise, then that particular theory has a counter example. If the theory is not fully compatible with our intuitions, the proponents of the theory should provide us with an explanation why our intuitions were wrong. If there is no plausible error theory available, then this can be seen as a further drawback for the theory of knowledge in question.

There are a few well known problems with the counter example method. First of all it relies heavily on our intuitions, and intuitions differ from person to person. They are seen as untrustworthy, unstable and subjective (Becker 2007, 3). I am inclined to accept this to a degree. But why am I then relying on intuitions? I do it for two reasons: (i) intuitions and behavior are the only clues we got in our search for an account of knowledge, and intuitions reflect our behavior and, (ii) the examples that I will present wake almost unanimous intuitions, which we have to take as truth telling. A plausible theory will explain why intuitions diverge in some cases while not in others.

The second problem with the counter example method is that it is unclear whether a single counter example to a theory is enough to undermine it. If the case rests on shifty intuitions or it is an extremely bizarre case, should it undermine the theory? I suggest not. Once one understands the method of building counter examples it is quite easy to come up with them. Probably no theory (even the ultimately correct one if there is such a thing) will be free of them all. All we can say is that the most plausible theory is the one which can deal with most of the cases and particularly the more general ones. Counter examples that feature wizards and what-not should not be taken too seriously. Instead the counter examples should be as mundane as possible.

Thirdly it is regrettable that the counter example method can have ill effects on ones writing. In worst cases the style is reduced to a catalogue of counter examples. I'll try to avoid that as best as possible. Keeping track of the counter examples can be tedious, and that is why I have devised an index featuring all the counter examples listing the first page where they are mentioned. All the counter examples are named and written in

capital letters the first time they occur. This will hopefully make it easier for the reader to follow the trail of the counter examples.

4.2. Possible worlds

As I said in the introduction, the reason why subjects seem to lack knowledge in Gettier cases is because their beliefs are only luckily true. A belief that is luckily true does not count as knowledge, even if the belief is justified. It seems plausible that beliefs that are true in virtue of luck would have easily been false. The reason why the subjects in Gettier cases lack knowledge is because they could have easily erred. Knowledge requires a certain level of stability and certainty. To know something is to be able to rule out various error possibilities. To know is to be in a state where one would not have easily erred.

'Would', 'could', 'easily' and their negations are modal terms. A sentence including a modal term describes how things could, would or might have been, or how things necessarily are or are not. They describe possible or impossible states of affairs. It seems that modal terms like these are central in explicating what knowledge really is; if the reason why the subjects in Gettier cases lack knowledge is that they would have easily erred in their beliefs, then some sort of modal condition could probably be immune to the Gettier cases. But the natural language terms 'would', 'might' etc. are vague. That is why I will use possible worlds as a tool for clarifying them. This approach has been favored by many other contemporary epistemologists as well (Becker 2007, Pritchard 2005, Sosa 1999, Hawthorne 2004, Greco 2012, Nozick 1981, Williamson 2000 and DeRose 1995). And even though the appeal to possible world semantics in explaining counterfactuals is controversial, it is considered to be a useful tool that clarifies the discussion (Williamson 2000, 150). But what are possible worlds?

We live in the actual world. The actual world is how things are. But many things might have been different. Even though I'm actually writing my thesis right now, I might very well have been reading instead. The possible world where I'm reading is a nearby one. It could easily have happened that I was reading right now. On the other hand, the

possible world where I'm on a vacation in Japan is a faraway possible world; much would have had to change for it to be the case that I was in Japan.

It is a common practice to order the possible worlds in terms of how similar they are to the actual world (Pritchard 2009, 26). Similarity is a vague concept and one could argue that the possible-worlds-talk only confuses the issue of how to handle subjunctive conditionals. Even though it is true that 'similarity' is vague concept at best, and context sensitive at worst, we still have a rudimentary understanding of which imagined possible worlds are more similar to the actual world than others (Becker 2007, 11). And even though similarity might be a vague term, it doesn't mean that it is ill-understood, and that it could not function as a primitive concept in an analysis (Lewis 1973, 91).

Because we use possible worlds to clarify modal notions and subjunctive conditionals, we have to lay down some truth conditions for modal statements. In terms of possible worlds a proposition is necessarily true if and only if it is true in all possible worlds. A proposition is necessarily false if and only if it is false in all possible worlds. A proposition is possibly true if and only if it is true in at least some possible world and possibly false if it is false in at least some possible world. A proposition is contingently true if and only if it is true in the actual world and false in at least some possible world. [1] A sentence featuring a subjunctive conditional like 'If *p* would have been the case, then *q* would have been the case' is true if and only if in the nearest possible worlds where *p* is true *q* is also true.[2]

This rough sketch of possible worlds should suit our purposes here. Although there are many issues with possible-worlds-talk we will brush them aside. As it often happens in philosophy, one will have to take some answers as granted to pursue another topic. Many discussions in ethics for example presuppose the possibility of free will, even though the metaphysical issue is far from settled (Pritchard 2009, 26). But even if it were later shown, that the appeal to possible worlds is unsustainable, the core of my project would be left intact. The reason for this is that we do not have to spell out the

[1] The actual world counts as a possible world in all of these definitions.
[2] Notice that this differs from the standard interpretation of the 'might' counterfactual given by Lewis. For Lewis a proposition of the form 'if *p* then *q*' is true if in the nearest possible world where *p* is true, *q* is also true (1973, 21). Remember also, that for Lewis the actual world counts as the nearest possible world. Our interpretation of the 'might' counterfactual does not focus on the nearest possible world where *p* is true, but on the group of the nearest *p* worlds. This modification to the Lewisian semantics has several benefits, as otherwise some of the modal conditions to be studied would end up being vacuously true if the subject's belief is true in the actual world.

subjunctive conditionals used in various definitions of knowledge in terms of possible worlds. We could use ordinary language just as well. The appeal to possible worlds just makes it easier to express the modalities involved.

5. Luck

If the anti-luck intuition is central to knowledge, then the necessary condition for knowledge that we aim to put forward has to exclude luck. But to be able to come up with such a condition we must first clarify what it is for some event to be lucky in the first place.

Winning a lottery is a paradigm example of a lucky event. What makes a lottery win a lucky event? First of all it is very rare to win a lottery. Chances are that you'll never win even if you gamble your whole life. There is also a price involved, so it is a significant event. Most people will never win, and those that do get a small fortune.

Pritchard characterizes the luck involved in a lottery win with the help of possible worlds. An event is lucky if it occurs in the actual world but not in a wide class of nearby possible worlds where the relevant initial conditions for the event remain the same as in the actual world (Pritchard 2007, 278). This captures nicely the rarity of a lucky event. Pritchard contrasts this with a paradigm case of a non-lucky event. If a skilled archer hits her target then the event is not lucky because in a wide class of nearby possible worlds where the archer shoots at the target she will continue to hit. In fact she would be unlucky if she didn't hit (2007, 278).

Another characteristic of lucky events is that they are significant to those whom they concern. Even an extremely rare event (one that occurs in the actual world but doesn't occur in any nearby possible worlds) will not count as a lucky event unless it is somehow significant to those whom it concerns. Think for example about a radioactive atom located on the far side of the moon that reaches half-life in a fraction of the time it would normally take. If this event has no particular consequences for anyone, then it

would be odd to say that it was a lucky event. Contrast this with an event that is rare and significant to someone. If I win a game of darts blindfolded then I would count myself as lucky. The reason for this is that winning pub-games is significant for me.

What is interesting about significance is that it is a neutral term in the sense that it doesn't explicate whether the consequences of the event are positive or negative. This plays well together with our account of luck, since there is such a thing as bad luck also. Slipping on a banana peel is unlucky because the consequences of the event are negative. Nonetheless the event is lucky in the relevant sense for our purposes. With these introductory remarks in play I will introduce the account of luck we will be relying on.

Luck: An event *e* is lucky for subject S just in case

(1) the event occurs in the actual world but not in a wide class of nearby possible worlds where the relevant initial conditions for the event remain the same

(2) the event is significant to S

This formulation draws heavily from Pritchard's definition, especially condition (i) (2005, 128). This kind of formulations has been put under pressure by Coffman and Riggs. They both argue for further refinements. Coffman demands that the subject must be a sentient creature and that the subject doesn't have control over the event in question (Coffman 2007, 396). Riggs argues that unless a lack-of-control-thesis is included in the definition of luck, we will not be able to accommodate the connection between true beliefs that are creditable to the agent and non-luckily true beliefs (Riggs 2007, 339).

I don't think that we have to take Coffman's first requirement seriously. Since we are interested in an account of luck that will allow us to illuminate the nature of knowledge, the requirement that the subject must be sentient is trivial. All subjects of knowledge are sentient. The requirement to include the lack-of-control-thesis is more severe, but on a closer inspection it is also redundant. Since we are interested about knowledge it is helpful to consider whether including the lack-of-control-thesis will have any ramifications to our judgments about whether a belief is luckily true or not. All of our

sensory beliefs are outside our direct control. We cannot voluntarily decide to believe in *p* unless it already seems to us that *p* is the case. Secondly condition (i) already exhausts the lack-of-control-thesis. If S could control whether the event *e* would obtain or not, then presumably the event would not only obtain in the actual world but also in the nearby possible worlds (where the relevant initial conditions for the event remain the same). So in a sense the lack-of-control-thesis is already incorporated in our analysis of luck. Furthermore since we are not attempting to give the sufficient conditions for a lucky event we can be satisfied with these two necessary conditions. [3]

6. Epistemic luck

Our next task is to apply our concept of luck to epistemological matters. Although knowledge is generally incompatible with luck, this is not always the case. If a detective happens to come by a crucial piece of evidence that shows that the defendant is guilty, her true belief based on her evidence will count as knowledge, even though she is lucky to have the evidence (Pritchard 2007, 279). The luck that is harmful for knowledge, targets the truth of the belief in question, not the fact that the subject has the belief. The detective was lucky to have the piece of evidence, but she was not lucky in her true belief that the defendant was guilty.

Let us now rephrase our definition of luck in terms of beliefs, not events.

Epistemic Luck: Subjects S's belief that *p*, is epistemically lucky just in case:

(1) *p* is true in the actual world

(2) *p* is false in a wide class of possible worlds where S believes that *p* while the relevant initial conditions for the truth of *p* remain the same as in the actual world.

Not surprisingly this formulation also draws heavily from Pritchard's definition of epistemic luck (2007, 280). Notice that it is not necessary to state explicitly that the belief has to be significant for the agent, since true beliefs are always significant to whoever holds them.

[3] For critique of modal accounts of luck see Lackey (2008). For defence against Lackey's arguments see Carter (2010).

Another point in favor for defining epistemic luck along the lines above is that it can accommodate the fact that not all instances of luck are incompatible with knowledge. Because we have defined epistemic luck in terms of the variation of the truth of the belief over a set of possible worlds, cases in which the agent is lucky to acquire the belief in the first place will not count as epistemically lucky. And since such cases do not undermine knowledge we can be quite satisfied with the result.

Now that we have spelled out the main motivation for modal epistemology and have an adequate definition of epistemic luck we can move forward to analyze modal conditions that aim to eliminate epistemic luck. The 'sensitivity' condition will be the first one to enter the fray.

7. Sensitivity

The following chapter will be devoted to the 'sensitivity' principle. First we will look at the virtues of 'sensitivity'. After that we will attempt to solve some of the problems facing 'sensitivity'. In the third subchapter we will briefly discuss some theories that incorporate 'sensitivity' as a part of other necessary conditions for knowledge. In the fourth subchapter we will pass our verdict regarding whether 'sensitivity' qualifies as a necessary and possibly sufficient condition for knowledge.

7.1. Virtues of 'sensitivity'

The aim of the following three sections is to consider whether 'sensitivity' - not-$p \rightarrow$ not-(S believes that p) - forms either a necessary or sufficient condition for knowledge. I will start by looking at the intuitions that support the sensitivity condition. After that I will consider the ramifications that accepting sensitivity as a necessary condition for knowledge has.

The sensitivity principle was first expressed by Nozick in *Philosophical Explanations* where he presents his tracking account of knowledge. The conditions for knowledge proposed by Nozick are the following:

(1) p is true

(2) S believes that p

(3) not-*p* → not-(S believes that *p*)

(4) *p* → S believes that *p*.[4]

(Nozick 1981, 172 & 176)

Condition (3) is commonly labeled as the sensitivity principle although in Nozick's terms a subject's belief is sensitive to the truth only if it satisfies all of the above conditions. Condition (4) is sometimes called the adherence condition (Zalabardo 2012, 42). The reason why (3) is considered to express the sensitivity principle is because most philosophers reject the adherence condition as a non-starter (Zalabardo 2012, 56), (Becker 2007, 44-47). Thus when I speak of the tracking account of knowledge I will be referring to Nozick's whole theory and when I speak about the 'sensitivity' condition I will be referring only the condition (3).

Nozick gives us a quick idea of what it is to know according to him: "To know that *p* is to be someone who would believe it if it were true [adherence], and wouldn't believe it if it were false [sensitivity]" (Nozick 1981, 178). The question is how we should interpret these subjunctive conditionals. Nozick uses possible worlds as a tool to clarify his conditions, but doesn't give us an explicit definition of what possible worlds count as relevant when judging whether a belief satisfies his conditions (1981, 176).

It is evident that conditions (1) and (2) concern the actual world which S inhabits. Both of them are necessary in all sensible accounts of knowledge, and thus they are under no kind of pressure. According to Nozick a subject's belief is such that "4 holds true if not only does he actually truly believe that *p*, but in the "close" worlds where *p* is true, he also believes it […] condition 3 speaks not of the whole not-*p* neighborhood of the actual world, but only of the first portion of it" (1981, 176). The definition of (4) in terms of possible worlds becomes quite clear in this paragraph. Nearby possible worlds where *p* is true are relevant for evaluating whether the condition is met. The definition of (3) however remains ambiguous. This is because it is unclear how "the first portion" should be interpreted. Does it refer to the nearest not-*p* world, or to the nearest group of not-*p* worlds? Greco goes with the former interpretation while for example Becker and Zalabardo go with the latter interpretation (Greco 2012, 195) (Becker 2007, 42) (Zalabardo 2012, 43). I will also go with the latter interpretation because it is a more

[4] '→' is to be understood as a subjunctive conditional.

charitable reading of the notion of sensitivity, and by adopting it one may dodge some bullets. Here is our interpretation of sensitivity made explicit:

Sensitivity: A subject S's belief that *p* is sensitive just in case

(1) in the nearest possible worlds where *p* is false, S does not believe that *p*.

Now we can test whether 'sensitivity' is fit to form the necessary, and perhaps sufficient, conditions for knowledge. To illuminate the virtues of the sensitivity condition I will test it against a few thought experiments. Let's start with a basic Gettier problem.

DUCK IN THE REEDS:

Peter is walking by a pond and sees a duck near the reeds on the other side. The duck swims into the reeds, so that Peter no longer sees it. Peter forms the justified true belief that there is a duck in the reeds. Unknown to Peter however the duck that he saw was not in fact a duck. It was a decoy used by hunters to entice other birds to the area. Luckily for Peter however there was a real duck in the reeds, so his belief is true.[5]

'Sensitivity' preforms admirably with respect to standard Gettier cases, such as the one above. Even though Peter has a justified true belief, his belief does not count as knowledge. Crucially the belief is not sensitive either. In the nearest possible worlds where there is no duck in the reeds, Peter continues to believe that there is a duck in the reeds. The sensitivity condition is able to handle the luck which undermines knowledge in standard Gettier cases.

BARN FAÇADE:

Henry is driving in the countryside with his son and identifies various objects for his edification. Henry sees a barn and tells his son "That's a barn". Henry has excellent eyesight and forms the corresponding justified true belief. Unbeknownst to Henry he has just entered a barn façade county. The area is scattered with convincing fake barns that would fool anyone. Intuitively however Henry doesn't know that he sees a barn. If he had happened to point at a mere façade he would still have believed that it was a barn. (Goldman 1976, 772-773)

[5] This case is analogical to Chisholm's 'sheep in the fields' case, though a bit more convincing. (Chisholm 1977, 105)

‘Sensitivity’ fares equally well with respect to more complicated cases, like the ‘barn façade’ case. Whereas in standard Gettier cases the subject’s belief is based on a false assumption or reached by a cognitive processes that errs, in the ‘barn façade’ case the subject’s cognitive apparatus is working properly. Henry’s belief is based on his actual visual experience and he gets the facts of the matter right.

In the standard Gettier cases the luck in play is intervening. If there hadn’t been a real duck in the reeds, then Peter would have formed a false belief. The luck fixes the subject’s false assumption. In the ‘barn façade’ case the luck in play does not intervene; it’s environmental. Henry really sees a barn, and does not perform any errors in his reasoning. Nonetheless, the epistemically inhospitable environment in which Henry is located ensures that his belief is only luckily true. (Pritchard 2010, 36)

Sensitivity condition can easily explain why Henry lacks knowledge. In the closest possible worlds where Henry’s actual belief is false, he will continue to believe in the false proposition. Presumably in those worlds Henry points a bit more to the left, where there is only a barn façade, and forms the same belief that he forms in the actual world. So Henry’s belief is not sensitive to the truth of the proposition which he believes in.

LOTTERY:

> Lottie and Luttie have each bought a ticket for a fair lottery with very long odds of a million to one. The Lottery has been drawn, but neither agent has heard the result. Lottie reflects on the fact that the odds are massively stacked against her and so, solely on this basis, forms the (true) belief that her ticket has not won. Luttie, in contrast, doesn’t even know what the odds for the lottery are, and certainly isn’t the sort of person to reflect on how these odds are stacked against her. But she reads the result of the lottery in a reliable newspaper, and so on this basis forms the (true) belief that her ticket has not won. (Pritchard 2012a, 177)

Intuitively Lottie lacks knowledge in this case while Luttie has knowledge. This might be surprising because Lottie’s probabilistic reasoning is sound. The odds are very much against her. Luttie’s belief on the other hand counts as knowledge, even though it could

be more probable that the newspaper misprinted the results, than for her to hold the winning ticket.

'Sensitivity' can explain this very elegantly. In the closest possible worlds where Lottie wins the lottery, she will still believe that she lost, based on her reflection on the extremely long odds. Lottie's belief is thus not sensitive. But in the closest possible worlds where Luttie wins the lottery, she will continue to read the lottery results from the same newspaper, and the newspaper will continue to report the results correctly in the *closest* possible worlds. Thus Luttie's belief is sensitive and the lottery case poses no problem for the sensitivity condition.

7.2. Problems for 'sensitivity'

Now I will turn to look at cases that pose some problems to the advocates of the 'sensitivity' condition. Let's start with the problems of belief-forming methods.

7.2.1. 'Sensitivity' and methods

Here is a case offered by Nozick, which shows that 'sensitivity' must be relativized to the actual belief-forming method used by the subject:

GRANDMA:

Granny the Grandma is especially good at judging whether a person is sick or not by looking. She visits her grandchild and sees that she is healthy. Granny forms the true belief that the child is healthy and she knows this. However, if the child had been sick her parents would not have allowed Granny to visit and would have lied about the child's medical condition in order not to upset the grandmother. (Nozick 1981, 179)

Intuitively Granny is in possession of knowledge in the above case but her belief is insensitive. In the closest possible worlds where her actual belief is false (where the child is sick), she will continue to believe that she is healthy, because the child's parents lie to her that she is healthy.

Examples like these prompted Nozick to relativize his account to the method actually used by the subject of knowledge. It should be easy to see that in the Grandma case the

Granny uses different methods of inquiry in the actual world and in the possible world where the child is sick. If we hold the method used by Granny constant and study only those possible worlds where she uses the same method, then her belief will end up being sensitive. In the closest possible world where the child is sick and Granny gets a glimpse of her, she will not believe that she is healthy. By relativizing the account to the actual method used, the advocates of sensitivity can dodge the problem above. The quick fix would look something like the following:

Sensitivity: A subject S's belief that *p*, which she came to believe via method M, is sensitive just in case

(1) in the nearest possible worlds where *p* is false, S does not believe that *p* via M.

This crude formulation might suffice, but there are several problems with it. There are cases where the subject's belief is overdetermined and where a method is one-sided, so that the subject would only believe *p* and never not-*p* by using that particular method. Here is a case of method overdetermination:

WISHFUL THINKING:

> A father believes that his son is innocent of a crime he is accused of. The father has no evidence, it is just a father's partiality. Later the father sees his son's innocence dramatically and unexpectedly proved in court. Beforehand the father merely believed that his son was innocent. Now he knows him to be innocent." (Armstrong 1973, 209)

Since in this case there are two methods at play, it is uncertain how the revised sensitivity condition should be applied. Wishful thinking is clearly not a method that is sensitive to truth, but relying on the courtrooms judgment should satisfy 'sensitivity'. But if the father has based his belief on the courtroom testimony, then his belief should probably count as knowledge. Intuitions might be shifty about this case, but that might just play in the hands of the proponent of 'sensitivity'. What might affect intuitions about this case is whether the father will continue to believe in his son's innocence regardless of the courtrooms verdict. If wishful thinking overrides the courtroom testimony, then we will probably want to deny that the father has knowledge.

Here is an example of a one-sided method by Zalabardo:

MEDICAL TEST:

> Consider, for example, a medical test for a condition with virtually no false positives but lots of false negatives. A positive result in the test virtually guarantees that the condition is present, but a negative result provides only very weak support for the hypothesis that the condition is absent, since lots of people with the condition test negative. (Zalabardo 2012, 58)

Zalabardo stipulates that a doctor who uses the above method of inquiry has knowledge and can know that a patient has the tested condition if the test results are positive. However a doctor using this one-sided method will never form a belief that a patient does not have the tested condition. The method is good only for affirming that a patient has the condition, not for determining that she does not have it. Because the method can only recommend a belief in *p*, the consequent of 'sensitivity' can never be satisfied. There is no possible world where S believes that not-*p* by using the same one sided method that she uses in the actual world. Luper-Foy has argued that Nozick's theory of knowledge must be restricted to two-sided methods, methods capable of recommending a belief both in *p* and in not-*p*, and this seems to be the case (Luper-Foy 1984, 28).

For reasons such as these Zalabardo rejects the move to relativize sensitivity to methods (2012, 62). In my opinion this creates even bigger problems, and Zalabardo is forced to admit that 'sensitivity' is not a necessary condition for knowledge. According to him 'sensitivity' is not a necessary condition for inductive and inferential knowledge, but we will return to these issues later. I want to point out that Zalabardo's rejection of 'method-sensitivity' is premature. Once we understand the relation between doxastic states and one-sided methods correctly, they will not pose a problem for 'sensitivity'.

Notice that the medical test warrants belief only in thinking that the patient has the condition. Supposedly if the test is any good, and if we aim to retain Zalabardo's stipulation that the doctor knows that a patient has the condition when the test results are positive, then the nearest possible worlds where the patient does not have the condition are such where the test does not indicate that she has the condition. In those possible worlds the test results are negative. However the negative test results don't warrant any belief. But in that case the doctor will presumably suspend her judgment about the patient's medical condition. But if that is the case, then the doctor's belief is

‘method sensitive’ after all. In the nearest possible worlds where *p* is false the doctor does not believe that *p*, she withholds her judgment, and this is compatible with ‘sensitivity’. Her belief is method sensitive after all.

Another question is how coarsely the methods should be individuated. Is it enough to determine what cognitive apparatus the subject was using, or should we aim for a more fine grained approach that takes into account the experiences based on which the subject forms her beliefs? This problem is analogous to the ‘generality problem’ for process reliabilism (Becker 2012, 91). The problem that is posed by the generality problem is how coarsely reliable processes should be individuated. The answers to this question will have clear consequences on the implications of any reliabilist theory (and a theory based around ‘sensitivity is a reliabilist theory in the relevant sense) and so the question must be resolved (Conee & Feldman 1998, 2). To see this consider the following case:

DACHSHUND:

> Oscar sees Dack the dachshund and believes there is a dog before him. If there weren’t a dog before him, a hyena would be there instead, which Oscar would misclassify as a dog. (Goldman 1983, 84)

If Oscar’s method is simply his use of vision, then his belief will end up being insensitive, even though Oscar knows that he sees a dog. This would indicate that methods must be defined more narrowly. Becker argues that the methods should be individuated as narrowly as possible (Becker 2012). The problem with this approach is that the satisfaction of ‘sensitivity’ will be trivial whenever one forms a true belief with a given method. Becker argues convincingly that this is not the case (Becker 2012, 97).

Even if we settle with Becker’s approach there remains the question whether we should define methods internally or externally. An internal characterization would be against our (and the sensitivity theorists) externalist agenda, so it would be better if we could come up with external criteria. Here is Nozick’s attempt to solve the current problem:

“A person can use a method (in my sense) without proceeding methodologically, and without knowledge of what method he is using. Usually, a method will have a final

upshot in experience, and then (a) no method without this upshot is the same method, and (b) any method experientially the same, the same "from the inside", will count as the same method." (Nozick 1981, 184-185)

Nozick has been criticized for the ambiguity at play here. On the other hand he says that the subject might not know which method he is using, and on the other hand he says that methods are individuated on how they "appear from the inside". The former is a nod to the externalist, while the latter is a nod to the internalist. Notice that Nozick cannot retort to a purely externalist definition, because then he would lose the skeptical consequences of his view, which he takes to be a virtue.[6] He has to appeal to the internalistic explanation. I will follow Nozick on this occasion, because otherwise we would end up with a twisted picture of the sensitivity condition.[7] Nonetheless I would rather individuate methods on purely external criteria.

This concludes our section of method relativity of sensitivity. Now I will turn to a more serious problem, the problem of inductive knowledge.

7.2.2. The problem of inductive knowledge

Here is a thought experiment introduced by Sosa which is supposed to show that 'sensitivity' is incompatible with many pieces of inductive knowledge that we take ourselves to know:

CHUTE:

Ernie is going to the elevator and on his way he drops his trash bag down the chute from his high rise condo. Presumably he knows that the bag will soon be in the basement. But what if the bag would not arrive to the basement? Then it would most probably have been snagged on the way down. This would be an incredibly rare occurrence. Ernie doesn't contemplate on such possibilities and as he releases the bag he forms the (true) belief that the bag soon reaches the basement. (Sosa 1999, 145)

[6] According to Nozick we cannot know the denials of skeptical hypotheses. Nozick saw this as a virtue for his view, but others found it another reason to abandon 'sensitivity' as a necessary condition for knowledge.
[7] Notice however that Tim Black (2002), who is an advocate of the sensitivity principle, defines methods on external basis.

It appears that Ernie does know that the bag will soon be in the basement. The reason why we judge that Ernie is in possession of knowledge is because it is extremely unlikely that Ernie's belief would be false. The closest possible world where the bag snags on the way down is a far away possible world. The bag is not going to snag on its way down. The snagging of the bag would be incredibly rare occurrence.

Notice however that this does not affect the fact that Ernie's belief is insensitive. In the closest possible worlds where his actual belief is false, he will continue to believe in the false proposition. It doesn't matter that the world is a very distant possible world. The sensitivity condition is at pains to accommodate inductive knowledge.

Here is another case:

PARKED CAR:

A few hours ago I parked my car on Liisankatu. Car thefts are extremely rare in this particular neighborhood, and from this I infer that my car is parked on Liisankatu. My belief is true and justified, but nevertheless insensitive. In the closest possible world where my car is not on Liisankatu, it has been stolen, and I still believe that my car is parked on Liisankatu.

'Sensitivity' would rob me of knowledge also in this case. This is contrary to intuitions, since we should be able to know many things based on reliable inductive inferences. The possible worlds where my belief is false are very distant, but this robs me of knowledge, according to the sensitivity condition. In fact any piece of inductive reasoning will be insensitive to the truth of the belief in question. This is a serious drawback for 'sensitivity'. Is there anyway the proponent of 'sensitivity' can save our inductive knowledge?

Becker's response is that we attribute knowledge all too liberally to agents who might be considered to have inductive knowledge. He would claim that I don't know that my car is parked on Liisankatu, I merely know that it's probable, and that Ernie doesn't know that the trash bag will soon be in the basement, he only knows it to be probable. (2007, 55-56)

Becker's reply isn't very satisfactory, since we would like to maintain that we have lots of inductive knowledge. Surely I know that the sun will rise tomorrow (based on

induction), even though my belief is insensitive! Zalabardo takes a different line of response to this problem, but it relies on his two-tiered conception of knowledge. As said above he does not think that 'sensitivity' is a necessary condition for knowledge in case of inductive and inferential knowledge. His probabilistic account can satisfy our intuition that we have inductive knowledge, but more on that later.

The sensitivity theorist could try to avoid the problem of inductive knowledge by reformulating the sensitivity condition. If she restricted the sphere of possible worlds that count as relevant when determining whether a belief is sensitive she would sidestep the problem. The revised sensitivity condition would look something like this:

Restricted sensitivity:

Subject S's true belief, that *p*, which she came to believe via method M, is restrictedly sensitive just in case within a restricted space of possible worlds (centered on the actual world)

(1) S does not believe, that *p*, via M in the nearest possible worlds where *p* is false.

(Greco 2012, 200).

This reformulation would allow the sensitivity theorists to accommodate our inductive knowledge. There are however ample reasons as to why the proponents of 'sensitivity' cannot afford to take this line of defense. 'Restricted sensitivity' is not compatible with the skeptical results that the advocates of 'sensitivity' are after. Recall that they want to uphold the 'intuitive' idea that we cannot know the denials of skeptical hypotheses. 'Restricted sensitivity' would make their account highly anti-skeptical, because if the relevant set of possible worlds include only nearby worlds, then our beliefs in the denials of skeptical hypotheses will automatically be sensitive, since possible worlds where skeptical hypotheses obtain are far away possible worlds, and thus outside the restricted space of possible worlds. Moreover 'restricted sensitivity' is very close to another competing modal condition, 'safety'. If 'sensitivity' is made extensionally equivalent to 'safety', then there is no point in advocating the more complicated 'restricted sensitivity' condition. As a last straw Greco demonstrates that 'restricted sensitivity' is inferior to 'safety' (2012, 201).

7.2.3. The problem regarding necessary truths

Inductive knowledge poses a severe problem at least for those philosophers who think that 'sensitivity' is a necessary condition for knowledge. Another crucial problem that faces the proponents of 'sensitivity' is the problem of necessary truths. A belief in a necessarily true proposition cannot be false, and thus such beliefs cannot fail to satisfy the sensitivity condition. The condition is trivially satisfied, so all beliefs in necessary true propositions should qualify as knowledge. But this cannot be the case, since clearly someone can believe in a proposition that is necessarily true, and yet fail to know it. Here is a case that illustrates the point:

MALFUNCTIONING CALCULATOR:

I use a calculator to count the product of 12 x 13 and as a consequence form the (necessarily) true belief that 12 X 13 = 156. Unbeknownst to me the calculator is malfunctioning, and generates answers at random.

Intuitively this is not a case of knowledge since one cannot use a malfunctioning calculator to gain knowledge. 'Sensitivity' however is trivially satisfied, because there is no possible world where my belief is false. That means that necessary truths pose an insurmountable problem for the proponents of 'sensitivity'.

I see two available lines of response. Either the 'sensitivity' theorist can restrict his theory of knowledge to contingently true proposition. Pritchard makes this move with respect to a different modal condition facing the same problem (Pritchard 2005). Alternatively the 'sensitivity' theorist can add another necessary condition for knowledge, which is not satisfied in cases like 'malfunctioning calculator'. Becker advances this strategy and demands that all beliefs that count as knowledge must also be reliably formed (Becker 2007). Whichever strategy one adopts, the consequences are not positive for the proponents of 'sensitivity'. Either the theory will be more complicated or it will not be the general theory of knowledge.

7.2.4. Violation of epistemic closure

The final problem that faces the proponents of 'sensitivity' is that 'sensitivity' violates epistemic closure. This is the most famous objection to 'sensitivity' and much has been written about it. Here is a rough sketch of the principle:

Closure:

Necessarily, if S knows that *p* and S knows that *p* entails *q*, then S knows that *q*. [8]

According to Williamson the idea behind closure is extremely plausible, since it articulates the intuition that "deduction is a way of extending one's knowledge" (2000, 117). On independent basis any philosopher should be inclined to accept a principle along the lines of 'closure'. The fact that 'sensitivity' violates 'closure' has been seen as an outright reason to reject any theory that presents 'sensitivity' as a necessary condition for knowledge.

The reason why 'sensitivity' violates epistemic closure is because a subject's belief in *p* might satisfy 'sensitivity', while the logically entailed belief in *q* doesn't satisfy the principle. These problems become highlighted when *p* is an everyday proposition (one that we take ourselves to know) and *q* is a skeptical proposition (one which we don't take ourselves to know). (Pritchard 2008, 7)

Here is an example: My belief that I'm sitting by my computer is sensitive, because were I not sitting by my computer I would not believe that I was. Thus this belief counts as knowledge. However the truth of this belief entails that I am not a brain in a vat (BIV), who merely thinks himself to be sitting by the computer. Nevertheless this latter belief is not sensitive and thus does not count as knowledge. If I were a BIV I would still believe that I was not. But according to 'closure' I should know that I'm not a BIV because this is logically entailed by my knowing that I'm currently seated in front of my computer.

Basically there are three ways to avoid the paradox. One can grant that we know the negations of skeptical hypotheses.[9] This line of response is not available for the

[8] Of course this formulation of the closure principle is inadequate. First of all in order for the subject to know that *q* she must make the inference from *p* to *q*, moreover the inference has to be valid. The subject must also retain her belief in *p* throughout the inference. If she came aware of a defeater to *p* and abandoned her belief in *p*, then her inference would not be made from known premises. Hawthorne (2004, 34) makes these requirements explicit, but nothing hangs on them here, so we can leave it as a footnote for now.

sensitivity theorist because the belief is not sensitive. The other option is to accept that we do not know many everyday propositions. Of course the skeptical consequences would be enormous, so it is understandable that almost no one is inclined to follow this strategy. Moreover since our beliefs in ordinary propositions often satisfy 'sensitivity', this line of response is not available for the proponents of 'sensitivity'. The third option is to deny 'closure', and for the sensitivity theorist this seems to be the only way out of the paradox. Becker (2007) for example takes this line of response. However there are others who have tried to reformulate 'sensitivity' so that it avoids the problem of closure (Murphy & Black 2012) and those who have challenged 'closure' and reformulated the principle so that it no longer poses so grave threat to 'sensitivity' (Bauman, 2012).

Even if the problem of closure is not a knockdown argument against 'sensitivity', it can still be seen as a factor that speaks strongly against it. Combined with the inevitable problems of necessary truths and especially inductive knowledge, the prospects of 'sensitivity' as a necessary or sufficient condition for knowledge are bleak. Surely a definition of knowledge that didn't have these drawbacks would be preferable.

7.3. Evaluating alternative theories based on 'sensitivity'

Now I will turn to briefly evaluate Zalabardo's proposal, according to which sensitivity is not a necessary condition, but still expresses an important aspect of knowledge. I will also look at Becker's approach who adds to 'sensitivity' another necessary condition. Finally we will consider DeRose's strategy of marrying the 'sensitivity' condition with contextualism and Nozick's good old tracking theory. We will start by examining Becker's proposal.

7.3.1. Becker's proposal

[9] This is a Moorean line of response and it is often put forward by those who advocate 'safety' as a necessary condition for knowledge. See Pritchard (2012b, 113) for a characterization of Mooreanism.

Becker claims that in addition to 'sensitivity' we need to have a 'reliability' condition in order to get the necessary and sufficient conditions for knowledge. Here is Becker's reliability condition:

Reliability:

"S's belief that *p* is formed by a belief-forming process or method[w] that produces a high ratio of true beliefs in the actual world and throughout close possible worlds." (Becker 2007, 88)

How does the addition of 'reliability' affect the prospects of a 'sensitivity' based account? It would seem that Becker's move allows him to deal with the problem of necessary truths. Recall the 'malfunctioning calculator' case? Intuitively my belief did not qualify as knowledge, because one cannot use a broken calculator to gain knowledge. My belief satisfied 'sensitivity', but it is easy to see that it does not satisfy 'reliability'. That is because the process I'm using (relying on a malfunctioning calculator) does not generate a high ratio of true beliefs in the nearby possible worlds. In a sense my belief was just luckily true, because in most nearby possible worlds I would have formed a different, yet false belief, concerning the question what is the product of 12 X 13.

So Becker manages to avoid one of the three big problems for 'sensitivity'. It is unable to deal with the problem of inductive knowledge because sensitivity is expressed as a necessary condition. Since sensitivity is never satisfied in cases of inductive knowledge, there is no inductive knowledge according to Becker's view. Closure will also be violated, but Becker doesn't see this as a too big problem (2007, 137). Others will be inclined to disagree.

7.3.2. Zalabardo's proposal

Zalabardo advances a disjunctive definition of knowledge. 'Sensitivity' is not a necessary condition for knowledge according to Zalabardo, but it nevertheless expresses something that is central about knowledge. It is a sufficient condition in cases of non-inferential knowledge. According to Zalabardo an agent has inferential knowledge either if she has adequate evidence for her belief or it is a standing belief (2012, 139). Adequate evidence for proposition H is defined as a high conditional probability for H

being true given the evidence E (p(H | E)). Zalabardo also introduces the following necessary condition for inferential knowledge in order to maintain the 'intuitive' idea that Moorean inferences don't count as knowledge:

"PI: S can have inferential knowledge of H based on the evidence provided by E only if S's belief in E confirms H." (2012, 98)

'Confirm' is supposed to be read here as increasing the likelihood of the believed proposition. The evidence that I have for my belief that I'm sitting by my computer does not increase the likelihood that I'm not a BIV because a BIV would have the same evidence if I was a BIV. My inference from 'I am sitting by my computer' to 'I am not a BIV', would not be an instance of inferential knowledge, even though I know the premiss. But from this follows that Zalabardo's account violates closure.

There are some advantages that Zalabardo's view has over standard views. He avoids the problem of inductive knowledge. Since 'sensitivity' has no sway over inferential knowledge, the fact that our beliefs based on induction are insensitive does not matter. The probabilistic account of adequate evidence and inductive reasoning should be able to deal with cases like 'chute' and 'parked car'.

Nonetheless there is another problem that faces Zalabardo. Recall the 'lottery' case ? In that case 'sensitivity' performed admirably and it could explain why Lottie had knowledge while Luttie did not have. The 'lottery' case was a case of inferential knowledge, but because Zalabardo has restricted 'sensitivity' in his account to non-inferential knowledge, the solution that 'sensitivity' would offer is not available for him. Worse still, his probabilistic account of inferential knowledge is unable to handle the case. Clearly Lottie has adequate evidence (in Zalabardo's terms anyway) in the case. The conditional probability that she is holding a losing ticket is extremely high. Thus there is a drawback in Zalabardo's approach compared with accounts that present 'sensitivity' as a necessary condition for knowledge.

Zalabardo's modifications do not save him from the problems caused by necessary truths. Furthermore Zalabardo's disjunctive account of knowledge is much more complicated than for example Becker's account. In my opinion Becker's account is preferable, because it can solve as many problems as Zalabardo's theory without the

added complexity. Neither account however is fully satisfactory, and I'm inclined to reject them both.

7.3.3. DeRose's proposal

Keith DeRose has attempted to marry the 'sensitivity' condition with contextualism. Though DeRose does not pursue a definition of propositional knowledge, but attempts only to solve skeptical problems, he has made some insights that help the proponents of 'sensitivity' to avoid the issues regarding epistemic closure and the problems of inductive knowledge.

DeRose's solution is contextualist at heart. According to DeRose some contexts (and skeptical contexts in particular) invoke higher standards of knowledge than others. The strength of ones epistemic position determines whether one continues to know things as the requirements of knowledge rise. DeRose gives us a rough sketch of the notion of the strength of epistemic position with the help of possible worlds:

> An important component of being in a strong epistemic position with respect to P is to have one's belief as to whether P is true match the fact of the matter as to whether P is true, not only in the actual world, but also at the worlds sufficiently close to the actual world. That is, one's belief should not only be true, but should be non-accidentally true, where this requires one's belief as to whether P is true to match the fact of the matter at nearby worlds. The further away one can get from the actual world, while still having it be the case that one's belief matches the fact at worlds that far away and closer, the stronger a position one is in with respect to P. (DeRose 1995, 34)

Importantly DeRose does not think that 'sensitivity' is a necessary condition for knowledge. According to him in some contexts we demand that a belief has to be sensitive to its truth in order to count as knowledge. Thus he introduces the following rule:

Rule of sensitivity:

"When it is asserted that some subject S knows (or does not know) some proposition P, the standards for knowledge (the standards for how good an epistemic position one must be in to count as knowing) tend to be raised, if need be, to such a level as to require S's belief in that particular P to be sensitive for it to count as knowledge." (DeRose 1995, 36)

It should be easy to see that DeRose's contextualist strategy allows him to sidestep the problems invoked by inductive knowledge. If 'sensitivity' is demanded only in contexts where the criteria of knowledge are very high, then it does not matter that our inductive knowledge is insensitive, at least in quotidian contexts. According to the contextualists it is indeed a virtue of the view that our ordinary knowledge disappears when we move into skeptical contexts.

To solve the problem of epistemic closure we must retort DeRose's notion of comparative strength of epistemic position. According to it any subject S is in at least as strong epistemic position with respect to *p* as she is with respect to *q* if it is the case that if S knows that *p*, then she knows that *q* and if she does not know that *q* she does not know that *p*. Moreover DeRose claims that any subject is in equally strong epistemic position with respect to propositions like "I am not a BIV" and "I have hands" regardless of how high or low the standards for knowledge are because a subject is in no better a position to know that she has hands than she is to know that she is not a BIV. Therefore if a subject does not know that she is not a BIV she does not know that she has hands. (DeRose 1995, 31-33)

Now it should be easy to see how DeRose solves the problem of epistemic closure: if my belief in *p* is sensitive, and it implies *q*, and I believe *q*, while my belief in *q* is insensitive, I can happily know that *q* because I happen to be in a context where the standards for knowledge are low. If the standards rise so high that they require beliefs to be sensitive, I no longer know that *q*. Crucially this does not undermine closure, since I no longer know that *p* either. Since my epistemic position regarding *p* and *q* was equally strong, it follows that if I do not know that *q*, then I do not know that *p*. *Voila*, closure is not violated.

Even though DeRose gives hope to the proponents of 'sensitivity' by solving the problem of closure and inductive knowledge, the help he offers comes at a steep price. First of all 'sensitivity' cannot be considered to be a necessary feature of knowledge.

Secondly, contextualism is not something that most epistemologists would welcome open heartedly. Thirdly, DeRose's strategy is ill-suited for anyone attempting to give a definition of knowledge. In fairness it must be emphasized that DeRose was not attempting to give a definition of propositional knowledge. Therefore though illuminating, DeRose's proposal can feature here only as an interesting insight on why 'sensitivity' seems to be crucial for knowledge.

7.3.4. The tracking account

Finally we will consider Nozick's tracking account of knowledge, which incorporates 'sensitivity' as a necessary condition for knowledge. Earlier we noted that Nozick's fourth condition for knowledge, 'adherence' was considered as a non-starter by some advocates of 'sensitivity' and decided to focus only on Nozick's third condition for knowledge. Now we will look at whether the fourth condition indeed is a non-starter, or does it allow the proponent of 'sensitivity' to dodge some bullets.

We should reformulate 'adherence' to reflect the modifications we made with respect to 'sensitivity'. The reformulated condition should be formulated as follows:

Adherence:

A subject S's belief that *p*, which she came to believe via method M, is adherent just in case:

(1) in the nearest possible worlds where *p* is true and S uses M to arrive at her belief whether *p*, S believes that *p*. (Nozick 1981, 179)

A belief that satisfies both 'sensitivity' and 'adherence' is said to 'track' the truth. According to Nozick the 'sensitivity' condition "tells us only half the story about how [a] belief is sensitive to the truth-value of *p*. It tells us how [a] belief is sensitive to *p's* falsity, but not how it is sensitive to *p's* truth." (Nozick 1981, 176) 'Adherence' tells us the latter part.

Of the problems that 'sensitivity' has encountered so far, the tracking account of knowledge allows the advocate of 'sensitivity' to deal with the problem of necessary truths. To see this, consider the 'malfunctioning calculator' case . Clearly there are nearby possible worlds where I do not believe that 12 x 13 = 156, while it is true, since

the calculator generates answers at random. I would have equally well have believed that 12 x 13 = 146, by relying on the same method. Thus my belief does not track the truth and I fall short of knowledge as our intuitions would have it.

There are also some other problems that are problematic for a proponent of robust 'sensitivity' theory that the tracking theorist can evade thanks to 'adherence'. For example the famous 'dictator' case presented originally by Harman (1973, 142-154) poses severe problems for 'sensitivity' but 'adherence' deals with the case satisfactorily. Be that as it may, many have argued that 'adherence' is too restrictive as a condition for knowledge. Here is a thought experiment introduced by Sosa in order to undermine 'adherence':

PELICAN:

When I see a pelican in my garden I ought to know that there is a bird in my garden. However, my belief that there is a bird in my garden is not adherent because it could easily have happened that there was no pelican but only a small robin in the bushes out of my sight. There will be close possible worlds where there is a bird in my garden, but I do not believe it because the pelican is no longer in my garden. Thus, my belief is not adherent to the fact that there is a bird in my garden. (Sosa 2009, 12)

As is clear from the example, Sosa claims that he has knowledge despite his belief is inadherent. However, a defender of Nozick could easily argue that only a crude form of 'adherence' is left unsatisfied by Sosa's belief, and that the more accurate method relativized 'adherence' condition is satisfied by Sosa's belief. Recall that Nozick defined methods not externally, as I image Sosa would do given his externalistic project, but more or less internalistically; "Usually, a method will have a final upshot in experience, and then (a) no method without this upshot is the same method, and (b) any method experientially the same, the same "from the inside", will count as the same method." (Nozick 1981, 184-185) A proponent of the tracking account could argue that Sosa would be using a different method in the counterfactual scenario where he does not see a pelican in the garden, since perception is a too broad category. Rather the method in play would be something more specific and probably related to birds in some way. Therefore it seems that the 'pelican' case does not pose any problems to the tracking theorist.

Becker has also argued that ‘adherence’ is too restrictive. He presents the following thought experiment originally introduced by Lehrer and Paxson:

GRABIT:

S sees Tom Grabit steal a book from the library and forms the true belief that Tom stole the book. However, Unbeknownst to S, Tom’s mother claims that Tom is out of town, and that his twin John Grabnot took the book, which in fact is a lie. (Lehrer & Paxson 1969, 228)

According to Nozick ‘sensitivity’ is met in this case, but it is unclear whether ‘adherence’ is met (1981, 191-192). Nozick sees ‘grabit’ as a borderline case of knowledge. Becker on the other hand argues that ‘adherence’ is not met, since he thinks that there are nearby possible worlds where S overhears what Grabit’s mother says and thus no longer believes that Tom took the book (2007, 46). Furthermore Becker argues that S knows that Tom took the book, but here we find conflicting intuitions. Lehrer and Paxson do not consider it a case of knowledge, unless the case is manipulated so that Mrs. Grabit is a pathological liar and that John Grabnot is a fiction of her demented mind (Lehrer & Paxson 1969, 228). Nozick sees it as a borderline case that is hard to box in. Nonetheless it should be clear that a proponent of the tracking account can reply to this case in the same way that she could reply to the ‘pelican’ case to accommodate Becker’s intuition. Since the world where S overhears what Mrs. Grabit says is one where S uses a different method to arrive at her belief regarding whether Tom took the book, that nearby possible world does not count as relevant when assessing whether ‘adherence’ is satisfied or not. But if that is the case then ‘adherence’ is met and S knows that Tom took the book.

It is difficult to deny that the tracking account of knowledge is stronger than a robust ‘sensitivity’ theory. It avoids the problem of necessary truths, whereas the barebones ‘sensitivity’ condition failed miserably. Sadly ‘adherence’ will not help the tracking theorist when it comes to the problems of induction and closure. Even the old tracking theory of knowledge is not wholly satisfactory. Nonetheless ‘adherence’ is far from a non-starter as a necessary condition for knowledge, as argued by Becker and Zalabardo.

7.4. The verdict on ‘sensitivity’

I have attempted to illuminate the prospects of 'sensitivity' mainly as a necessary condition for knowledge. It seems fair to say that 'sensitivity' doesn't express a necessary condition for knowledge, at least for all kinds of knowledge. I've also demonstrated that 'sensitivity' cannot be a sufficient condition for knowledge. We can have sensitively true beliefs that do not count as knowledge.

My main goal was to accommodate the anti-luck intuition with the sensitivity principle. It proved out that 'sensitivity' does not cash out this intuition. The reason why 'sensitivity' failed this task is relatively simple. When we are considering whether a belief is luckily true, we are interested only about the possible worlds that are near the actual world. If the belief is true in the actual world but not true in most of the nearby possible worlds, then the belief is luckily true. 'Sensitivity' was not restricted to just the nearby possible worlds. When evaluating whether the sensitivity condition was met we had to traverse to the nearest possible worlds where the proposition believed in the actual world was false. Depending on the proposition in question those possible worlds could be nearby possible worlds or very distant ones, and even the distant ones were relevant. But why should we care about the distant possible worlds? Our account of luck is not concerned with them; only nearby possible worlds mattered. If we are after an anti-luck condition we should not care about all too distant possible worlds. The possible scenarios they present are not relevant when determining whether we have knowledge. 'Sensitivity' made them relevant alternatives, so 'sensitivity' as a condition for knowledge should be abandoned.

In the next chapter we will focus on 'safety' which is another modal condition. 'Safety' is not concerned with the faraway possible worlds, so it might just be the condition we are looking for.

8. Safety

The following chapter will focus on examining another modal condition; 'safety'. First we will give a preliminary account of the 'safety' principle and look how it deals with some basic thought experiments. In section 8.2. will proceed to examine cases that demand modifications to our formulation of 'safety' or pose insurmountable problems to the advocates of 'safety'. Finally in section 8.3. we will examine alternative formulations of the 'safety' principle that have been offered in literature.

8.1. ’Safety’ demonstrated

The guiding idea behind safety is that in order to know, one has to be safe from error. If you know something then your belief could not easily have been false. According to the advocates of ‘safety’ S knows that *p* if and only if:

(1) *p* is true
(2) S believes that *p*
(3) S believes that $p \rightarrow p$

The two first conditions are trivial and therefore under no pressure. The third condition expresses the ‘safety’ condition. In terms of possible worlds it could be read as

Strong safety: S’s true belief that *p* is safe if and only if

(1) In all the nearby possible worlds where S continues to believe that *p*, *p* remains true.[10]

The difference between ‘strong safety’ and ‘sensitivity’ is that they hold different possible worlds as relevant when determining whether a subject has knowledge. According to ‘sensitivity’ the relevant possible worlds were the nearest possible worlds where the proposition believed in the actual world was false. The condition was satisfied if the agent ‘tracked’ the truth and did not believe in the false proposition. Depending on the modal nature of the proposition in question the worlds that counted as relevant for ‘sensitivity’ could be either nearby possible worlds or ones that were very far away from the actual world. On the other hand the possible worlds that counts as relevant when determining whether a subject’s belief is safe are never far away possible worlds. Only the nearby possible worlds where the agent continues to form the same belief as she forms in the actual world count as relevant. If the subject’s belief remains true in all the nearby possible worlds then her belief is safe.

Given this fact it would seem that ‘strong safety’ is much better suited to accommodate the anti-luck intuition which has a central role in our analysis of knowledge. After all when we are trying to eliminate luckily true beliefs we are interested only in the nearby possible worlds. Recall that a belief was luckily true only if it was true in the actual

[10] It is more fruitful to start out with a robust reading of the safety principle. Later we will look at weaker readings of ‘safety’.

world but false in most of the nearby possible worlds. ‘Strong safety’ is obviously analogues with ‘epistemic luck’ (Carter 2010, 520). In fact what ‘strong safety says is that beliefs that are lucky in the sense of ‘epistemic luck’ should not count as knowledge and beliefs that do not count as lucky according to ‘epistemic luck’ do count as knowledge. This feature makes ‘strong safety’ symmetrical with ‘epistemic luck’. ‘Sensitivity’ on the other hand was asymmetrical with ‘epistemic luck’. It made the faraway possible worlds as relevant, and thus failed. Therefore ‘strong safety’ might triumph where ‘sensitivity’ missed the mark.

To demonstrate the virtues of ‘strong safety’ let us pitch it against some of the same thought experiments that we tested on ‘sensitivity’. ‘Strong safety’ performs admirably with respect to standard Gettier cases. Recall the ‘duck in the reeds’ case ? Peter’s true belief that there is a duck in the reeds is not safe, because there is a nearby possible world where Peter believes that there is a duck in the reeds but there isn’t. It could very easily have been the case that there was no duck in the reeds. Peter’s belief is only luckily true and ‘strong safety’ can accommodate this fact.

Next in line is the ‘barn façade’ case . Henry’s belief in this case is not safe since there are nearby possible worlds where he continues to believe that ‘that thing there is a barn’ even though his belief is false. He could easily have pointed a bit more to the left, where there is only a barn façade, and he would have been fooled by the façade and believed that it was a barn. ‘Strong safety’ can deal with environmental luck as well as with intervening luck.

The ‘lottery’ case is a bit more trickier. Recall that in this case the intuitive judgment was that Lottie lacked knowledge while Luttie had knowledge. ‘Strong safety’ can deal with this case because apparently winning a lottery is an event that occurs in nearby possible worlds (at least if you own a ticket). The reason for this is that “all that needs to be different, after all, is that a few coloured balls fall in a different configuration” (Pritchard 2012a, 178). Since Lottie will win the lottery in a nearby possible world there is a nearby possible world where she believes falsely that she lost the lottery based on her probabilistic reasoning. Lottie’s belief is not safe, and therefore not knowledge. Luttie’s belief on the other hand is safe. She acquires her belief that she has lost the lottery by reading the results in reliable newspaper. In the nearest possible worlds where she won the lottery the newspaper will report the results correctly, and thus she will not

believe that she lost. ‘Strong safety’ gives the correct verdict regarding the ‘lottery’ case.

The above cases were such where ‘sensitivity’ fared as well as ‘strong safety’. Now we shall turn to examine some cases where ‘strong safety’ preforms better. Recall the problem of induction and especially the ‘chute’ case ? ‘Sensitivity’ could not deal with cases featuring induction, but ‘safety’ can. Intuitively Ernie acquired knowledge in the ‘chute’ case. ‘Strong safety’ can accommodate this result since presumably, if we are to maintain the intuition that Ernie has knowledge, there cannot be nearby possible worlds where the bag snags on the way down. The bag gets snagged in the chute only in faraway possible worlds since the snagging of the bag would be an extremely rare event. Because the bag does not snag in the nearby possible worlds, there will not be any nearby possible worlds where Ernie believes falsely that the bag will not soon be in the basement. ‘Strong safety’ gets the right result in cases of inductive knowledge.

‘Strong safety’ can also respect epistemic closure, whereas ‘sensitivity’ could not. According to ‘strong safety’ if I know that I have hands I can deduce from this, and come to know, that I am not a handless BIV who just thinks that he has hands. The reason why ‘strong safety’ can accommodate epistemic closure is because the possible worlds where the skeptical hypotheses obtain are faraway possible worlds. In all the nearby possible worlds where I believe that I have hands I really do have hands. It does not matter that there are some faraway possible worlds where I am a handless BIV, because faraway possible worlds are never relevant when determining whether a belief is safe or not. Clearly such anti-skeptical results are very welcomed. What is problematic for the proponents of ‘strong safety’ is to explain how we can know the denials of skeptical hypotheses so easily, but more on that later. Now we will turn to look at some cases that are problematic to the safety principle.

8.2. ‘Safety’ under threat

In this chapter we will turn to look at cases that threaten ‘strong safety’. Some of the cases require us to refine our definition while others will guide us to understand it in the correct way. Finally some of the problems presented might be insurmountable and require that we reconsider whether ‘safety’ can be seen as a necessary condition for knowledge.

8.2.1. Greco's dilemma

Actually we have already presented two thought experiments that taken together threaten to tear 'strong safety' apart. Although it seemed that 'strong safety' is adequate to deal with both the 'lottery' and the 'chute' case this is far from clear. The problem is that, if 'strong safety' is able to deal with the 'lottery' case, it will be unable to deal with cases of inductive knowledge, like the 'chute' case. Greco was the first to present this problem, and we will accordingly call it 'Greco's dilemma' (2007, 300-301). Here is the problem:

In the 'lottery' case most of the possible worlds where Lottie continues to believe that she has lost the lottery she will have lost it. The number of very close nearby possible worlds where her belief turns out false is minimal. Nonetheless this is enough to rob her of knowledge. If the safety principle is to be able to accommodate this result it has to be formulated so that the belief the agent formed in the actual world remains true in all the nearby possible worlds where she continues to hold the belief (this is the formulation of 'strong safety' as we've given it). Unfortunately this strong version of the safety principle is unable to deal with the 'chute' case. Presumably there can be a miniscule number of nearby possible worlds where the bag gets snagged in the garbage chute, because there is a minor imperfection in the shaft. But even if there were a small amount of nearby possible worlds where the bag gets snagged in the shaft, we would still attribute knowledge to Ernie.

The 'strong safety' is unable to cope with the 'chute' case since it demands that the subject's belief, that *p*, has to remain true in *all* the nearby possible worlds where the subject believes that *p*. Perhaps a weaker version of the safety principle would be able to deal with both cases? Consider 'weak safety':

Weak safety: S's true belief that *p* is safe if and only if

(1) In most nearby possible worlds where S continues to believe that *p*, *p* continues to be true.

'Weak safety' is able to deal with the 'chute' case. However it should be obvious that it is not able to deal with the 'lottery' case, since Lottie's belief satisfies it. Lottie's belief continues to be true in most nearby possible worlds where she holds it, but nevertheless

she lacks knowledge. Weakening the safety principle does not seem to solve the problem.

Becker has argued that there might not be any coherent version of the safety principle which would bypass Greco's dilemma (Becker 2007, 65-67). That claim is not well established, but it is quite certain that Greco's dilemma demands that the safety principle must be refined in some way. Luckily Pritchard has offered us a neat solution out of this dilemma.

To see the solution, consider the scope of possible worlds, where Lottie's belief that she has lost the lottery is false. Presumably there are only a few possible worlds where her belief is false, but these worlds are very close to the actual world. She could very easily have won, since only a few balls would have to have fallen in a different configuration. Now consider the scope of possible worlds where Ernie's belief that the trash bag will soon be in the basement is false. If we are to maintain the intuition that Ernie has knowledge, then there can be only a few possible worlds where his belief is false. Furthermore these possible worlds cannot be the very closest of the nearby possible worlds. Otherwise Ernie's belief would have been very easily false, and if that is the case, then we would not attribute knowledge to him. (Pritchard 2007, 292-293)

Once we understand this we can formulate the safety principle so that it gives greater weight to the very closest nearby possible worlds, and less weight to the worlds that are a bit further away. (Pritchard 2007, 292)

Safety: S's true belief that *p* is safe if and only if

(1) in nearly all nearby possible worlds where S continues to believe that *p*, S's belief is true,
(2) and in all of the very closest nearby possible worlds where S continues to believe that *p*, S's belief is true.

'Safety' is now able to deal with both the 'lottery' case and the 'chute' case. Both (1) and (2) are satisfied in the 'chute' case, while Lottie's belief does not satisfy (2) in the 'lottery' case.

Notice that while the strong version of safety was analogues with our analysis of a luckily true belief, the new revised 'safety' is not. Either there is something amiss with

'safety' or our characterization of luckily true beliefs. It seems plausible that we should modify our account of luck. This is not an *ad hoc* solution since there are independent grounds for thinking that the very nearest possible worlds should carry more weight when determining whether and event or belief is lucky or not. If a sharpshooter misses me by a meter I am lucky to be alive. If he misses me by 10 centimeters I am even luckier to be alive. The reason for this is that the world where I die in the latter example is much closer to the actual world than the world where I die in the former example. The very nearest possible worlds should weight more than the worlds that are further off. Our revised account of a lucky event is thus the following:

Luck*: An event *e* is lucky for subject S just in case

(1) The event occurs in the actual world but not in most of the nearby possible worlds, nor in any of the very closest possible worlds, where the relevant initial conditions for the event remain the same as in the actual world
(2) the event is significant to S

Similarly we should reformulate our definition of a luckily true belief;

Epistemic luck*: Subject S's belief that *p*, is epistemically lucky just in case:

(1) *p* is true in the actual world
(2) *p* is false in a wide class of possible worlds, and in the majority of the very closest possible worlds, where S believes that *p*, while the relevant initial conditions for the truth of *p* remain the same as in the actual world. [11]

8.2.2. Problems with closure

Earlier we noted that 'safety' can preserve epistemic closure, and it would seem that it can. Alspector-Kelly has however argued to the contrary (2010). According to the closure principle if S knows that *p* and S knows that *p* entails *q*, then S knows that *q*. If S knows that *p*, then S's belief in *p* has to be safe. And if *q* follows from *p* and S infers it from p, then S's belief in *q* must also be safe. But in that case closure is preserved! Alspector-Kelly argues however that closure is violated if there is a possible world

[11] Carter gives a similar redefinition of epistemic luck in his 2010, 523. The difference is that he requires that the belief has to be false in at least *some* of the very closest possible worlds, whereas I require that is has to be false in most of them, if not in all of them.

where *q* is false and S believes it not on basis of *p* (which is false and not believed by S in this particular possible world) but on some other basis *r* (2010, 134).

This problem is however illusory once we understand 'safety' correctly. It seems obvious that the method by which S comes to believe *q* in nearby possible worlds must be kept constant. We noticed that 'sensitivity' required relativization to methods and 'safety' is no different in this respect. Once we have relativized 'safety' to the method actually used by the subject it will preserve closure.

8.2.3. Method relativization

The need to relativize 'safety' to the method used in the actual world became apparent in the last subchapter and the issue will be settled now. Without modification 'safety' will give the wrong verdict on the 'grandma' case . There is a wide range of nearby possible worlds where Granny believes that her grandchild is healthy, even though she is sick. Crucially in these worlds Granny relies on the child's parent's testimony, not on her reliable vision. If we keep the method constant across the nearby possible worlds that count as relevant when determining whether the subject's belief is safe, then Granny will no longer falsely believe that her grandchild is sick in any relevant nearby possible world. The revised safety condition can be formulated as follows:

Safety: S's true belief that *p* is safe if and only if

(1) in nearly all nearby possible worlds where S continues to believe that *p* (by the same method M that she uses in the actual world), S's belief is true,
(2) and in all of the very closest nearby possible worlds where S continues to believe that *p* (by the same method M that she uses in the actual world), S's belief is true.[12]

Recall that the 'sensitivity' condition faced a multitude of problems regarding method relativization. Proponents of 'safety' however can sidestep most of these problems. In 'wishful thinking' the father believed his son to be innocent of a crime based purely on his partiality. Later the son's innocence is proved at court, after which the father knows that his son is innocent. 'Safety' can deal with this case because wishful thinking does

[12] From here on all references to 'safety' will refer to this method relativized version of the condition. Exceptions will be duly noted.

not constitute a safe method while relying on the courtroom verdict does. By relying on wishful thinking the father will believe his son to be innocent in all possible worlds, even in those nearby possible worlds were the son is guilty. Thus he does not satisfy 'safety'. Later he uses a different method, which is safe. Supposedly there are no (or nearly no) possible worlds where the father believes his son to be innocent by relying on the courtroom verdict, while the son is guilty. However if wishful thinking and the courtroom testimony conflict and the father still relies on wishful thinking, then there will be a fairly large set of possible worlds where the father believes his son to be innocent while he is in fact guilty, and in that case 'safety' is not satisfied. 'Safety' is able to accommodate the intuitive judgments regarding 'wishful thinking'.

It should be easy to see that one-sided methods do not pose any problems to the advocates of 'safety'. The 'medical test' case was problematic to 'sensitivity' because apparently the consequent of 'sensitivity' could never be satisfied in the case. That was because a one-sided method can only recommend a belief in *p* and never in not-*p*. As we saw, the proponents of 'sensitivity' could dodge this problem, but it is important to note that the problem does not even rise to 'safety'. This is because 'safety' is concerned only with the situations where the subject believes that *p*. What matters is whether *p* is true in the nearby possible worlds where the subject believes that *p*. 'Safety' can deal with the 'medical test' case without any problems.

There are however two genuine problems that face 'safety' regarding method relativization; the generality problem and the problem of whether to individuate methods internally or externally. The proponents of 'safety' can appeal to the same solutions as the advocates of 'sensitivity', at least regarding the generality problem. I do not aim to solve the generality problem here. The reader must be contended with Becker's solution to the problem, according to which the methods must be individuated as narrowly as possible but still so generally that they can be applied in different situations (2007, 51).[13]

According to Becker methods M' and M* are the same method only if they are experientially the same (2007, 51). He has to make this acknowledgement to the skeptic and the internalist because he wants to be able to accommodate the intuitive judgment

[13] This solution to the generality problem allows us to deal with the issues that Alspector-Kelly raises towards the safe-basis account regarding epistemic closure. Alspector-Kelly admits this (2010, 136-137).

that we cannot know the denials of skeptical hypotheses. The proponents of 'safety' have no wishes to make this kind of acknowledgements to the skeptic. In fact many of the 'safety' theorists consider themselves as neo-mooreans. According to them we do know the denials of skeptical hypotheses.[14] Thus the proponents of 'safety' can happily retort to the more explicit way of defining methods externally. According to this line of reasoning, the cognitive methods M' and M* could be defined as the same method for example only if they arise from similar neural activity and similar kind of (external) stimulus.[15] Exactly how the cognitive methods on external basis are to be defined should be left to cognitive and neural scientists.

Furthermore, defining methods externally has the added advantage of strengthening our position against the skeptic. Even if there was some nearby possible worlds where I was a BIV I would still know that I am not. This is because if we hold the belief-formation method constant across the relevant possible worlds where I believe that I have hands, then all the nearby possible worlds where I use my actual belief formation method are such where I have hands. If methods are defined externally, the actual method used by me will not be the same method that I use in the BIV world. The BIV does not have the same kind of stimulus, mainly because it lacks the same kind of perceptual apparatus that I possess.

8.2.4. Methodological issues regarding the space of possible worlds

The proponents of 'sensitivity' did not have any considerable problems regarding the application of possible world semantics. The similarity ordering of possible worlds was enough to them, since 'sensitivity' demanded that a subject does not believe in the proposition believed in the actual world in the *nearest* possible world where the said proposition is false. The nearest possible world, where the proposition believed in the

[14] Later we will look at the problem of easy knowledge generated by the neo-moorean line of reasoning.
[15] Of course two methods M' and M* that would count as the same method from an external point of view tend to count as the same method from an internal point of view as well. Presumably similar neural activity gives raise to similar experiences. Nonetheless experiential indistinguishability of M' and M* will not guarantee that they will count as the same method from an external point of view.

actual world was false, is the not-*p* world which is most similar to the actual world. The sensitivity theorist does not have to add anything to the possible world semantics.

Unfortunately the advocates 'safety' need to say something more about the space of possible worlds. That is because the definition of 'safety' includes the notion of 'nearby possible worlds'. But what counts as a nearby possible world? In section 4.2. I gave an idea of what I mean by nearby and faraway possible worlds, but it is fair to say that the picture drawn there is too vague to be a crucial feature of the definition of knowledge. Alspector-Kelly for example voices a complaint about this issue regarding the safety principle: "given the dramatic shift in the requirements for knowledge […] when the boundary between worlds nearby and far is crossed, one would expect a clear, non-arbitrary specification of the line separating them." (2010, 132). Regrettably I have to admit that I do not have a better characterization available. Perhaps the border between nearby and faraway possible worlds is vague, and it cannot be determined, or it shifts with the context of knowledge attribution. DeRose has pursued this context sensitive interpretation of the nearby possible worlds, and if one has contextualist intuitions, then one could use this backdoor to bring contextualsim into an otherwise modal theory of knowledge (1995).

8.2.5. Four fatal problems

So far 'safety' has eluded every counter example. A theory incorporating 'safety' as a necessary or even sufficient condition for knowledge would be extremely powerful. Alas, 'safety' is bound to fail regarding cases featuring necessary truths. Recall the 'malfunctioning calculator' case ? Clearly I lack knowledge in the case. One cannot use a broken calculator to gain knowledge. Unfortunately for the proponents of 'safety', my belief is safe. There is no nearby possible world where I hold the belief that 12 x 13 = 156 while the belief is false, for the simple reason that there is no possible world where 12 x 13 = 156 is false. 'Safety' has a clear cut counter example, and its defenders should have something to say about cases featuring necessary truths.

There are a few strategies available to the advocates of 'safety'. One can restrict 'safety', and attempt to give a theory of knowledge featuring only contingent propositions as done by Pritchard in his 2005. Alternatively one can add other necessary conditions for knowledge that are not fulfilled in cases like 'malfunctioning calculator'.

The dangers of this maneuver are that it might be an *ad hoc* solution. One would need good, theory independent, grounds for bringing additional conditions to the fore. The third option is just to bite the bullet. Of course this course of action is never without drawbacks, and as will soon become apparent, it would lead down a slippery slope. The reason for this being, that there are counter examples that have the same structure that 'malfunctioning calculator' has, even though they do not feature necessary truths. This naturally undermines the move to restrict 'safety' to contingent propositions also. To see this consider the following case presented by Greco:

FROG:

John sees a frog, takes it to be green and as a consequence believes it to be green. In fact all the frogs in the environment in which John finds himself are green due to evolutionary mechanisms. John however does not know this, and more over he happens to be colorblind. John could easily be wrong about many things being green, but not about frogs, because all frogs are green. (Greco 2007, 301-302)

John's belief that the frog is green is safe, since there are no nearby possible worlds where he believes it to be green, while it is not. Intuitively John does not know that the frog is green, because he cannot distinguish between different colors. Notice also that John's belief is not necessarily, but only contingently, true. There are some faraway possible worlds, where the frog is of a different color. Thus restricting 'safety' to purely contingent truths does not save the theory from counter examples. Additionally, if one is willing to bite the bullet regarding necessary truths one should for consistency's sake bite the bullet in cases like 'frog' also. This is would render the theory pretty useless, so we should focus on trying to marry 'safety' with some other condition that is able to rule out cases like 'malfunctioning calculator' and 'frog'.

Before we do that however, we need to look at two problems that are also fatal for 'safety'. The first problem is related to the 'frog' case. Notice that in the 'frog' case the relation between the truth of the belief and the formation of belief is incidental. John just happened to form the belief that the frog is green, he could have taken it to be it red as well, but he did not. Luckily all the frogs happened to be green. The next case is a mirror image of the 'frog' case.

BRAIN LESION:

Alvin has a brain lesion. A peculiar fact about Alvin's brain lesion is that it causes its bearer to form the belief that she suffers from a brain lesion. Accordingly Alvin believes that he suffers from a brain lesion. (Pritchard 2009, 54)

Clearly Alvin does not know that he has a brain lesion. Nonetheless his belief is true, and it satisfies 'safety'. There are no nearby possible worlds where Alvin believes that he suffers from a brain lesion while his belief is false.

This case can be seen as a mirror image of the 'frog' case because Alvin really does have a cognitive capacity that produces reliably true beliefs. What hinders us from granting Alvin with knowledge is that the process by which he comes to form the belief seems overly strange.

I find my intuitions about this case a bit shifty, mainly because humans tend to use quite bizarre methods of inquiry. The autistic twins, who can spontaneously generate twenty digit prime numbers in Oliver Sacks book *The Man Who Mistook His Wife for a Hat and Other Clinical Tales* is particularly telling example of knowledge generation through extraordinary cognitive processes.

Let us move forward to the last fatal problem facing 'safety'. Recall the 'barn façade' case? 'Safety' gave the right verdict regarding that case, but with just a slight alteration the case will cause serious trouble to the proponents of 'safety'. Here is Kripke's red barn case adopted from Cohen:

RED BARN:

> Suppose there is a region where there exist barns along with barn replicas visually indistinguishable from actual barns. The residents of the region picked out all the sites and at each one flipped a coin to determine whether they would put up a real barn or a replica. As it turns out all the replicas are green. S is unaware of the replicas and seeing an actual red barn, comes to believe that there is a red barn before him. (Cohen 2004, 20)

This case causes two kinds of problems for the proponents of 'safety'. According to 'safety' the subject knows in the above case that there is a red barn in front of her. There are no nearby possible worlds where she believes that there is a *red* barn in front of her while her belief is false, so she satisfies 'safety'. This is already problematic, since in

the original ‘barn façade’ case the subject would seem to lack knowledge. What is even more severe is that the subject in the ‘red barn’ case does not know that there is a barn in her plain view; there is a number of nearby possible worlds where she believes that there is a barn in front of her, while there is mere façade in front of her. Thus she can know that there is a *red barn* in front of her while not knowing that there is a *barn* in front of her! But the latter is logically entailed by the former! It would seem that ‘safety’ violates closure after all!

The solution that the advocates of ‘safety’ need, to solve the problems caused by the ‘red barn’ case, has to entail that the subject does not know that there is a red barn in front of her. That is the only route to preserve closure and keep the intuitive response to the case. Notice, that in order to preserve closure, it would be enough to revise ‘safety’ so that the subject would know that there is a barn in front of her. That response would however be counterintuitive. It is a widely accepted that the subjects of barn façade cases lack knowledge. To give a contradictory answer would require a good error theory, and those are in short supply.

With these four fatal problems in mind we can turn to look at modifications of ‘safety’. Can ‘safety’ be elegantly joined with other conditions of knowledge, or be redefined, in order to yield optimal results regarding the problematic cases?

8.3. Reformulations of ‘Safety’

In this section we will be looking at reformulations of ‘safety’. First we will take a look at safe indication accounts of safety, proposed by Luper and Sosa. After that we will examine a modal theory of knowledge that conjoins ‘safety’ with a virtue-theoretic condition.

8.3.1. Safe indication

Steven Luper (2006) and Ernest Sosa (2004) have defended a safe basis or safe indication account of the safety principle. Here is Luper’s formulation of the safe indication account:

Safe indication: at time t, subject S knows *p* by arriving at the belief *p* through some method M only if:

(1) M would, at t, indicate that *p* was true only if *p* were true. (Luper 2006, 162)

Here is Sosa's version:

Safe basis: S's belief that *p* is safe if and only if

(1) S holds it on a basis that his belief would not have without being true. (Sosa 2004, 292)

The key difference between Luper's and Sosa's formulation is between the terms 'basis' and 'method'. I take it that 'method' means something similar that we meant by methods when we relativized the safety principle to the actual method used. Methods are tools and ways of inquiry, like perception, deduction, induction, intuition and memory. But Sosa does not mean this by 'basis'. 'Basis' refers to the deliverances of our cognitive abilities. A deliverance is a "state of affairs wherein something seems to oneself to be so - a more or less complex state that may feature intellectual as well as sensory content." (Sosa, 2009, 13)

Sosa's formulation relies heavily on the experience on which the belief is based, while Luper focuses on the reliability of the method used by the subject of knowledge. This divergence has a critical impact on how the 'safe basis' and 'safe indication' accounts handle the problematic cases presented in the last section. Now let us pitch them against the alleged counterexamples. We shall first look at the 'malfunctioning calculator' case.

It should be easy to see that the 'safe indication' account can evade the problem caused by the 'malfunctioning calculator' case. A malfunctioning calculator could indicate anything regardless of the truth of the indicated proposition. Therefore my belief would not be safely indicated. The 'safe basis' account on the other hand seems to be unable to deal with the case. The reason for this is that the deliverance of it appearing to me that 12 x 13 = 156 could not indicate something that is false, because the proposition believed is true out of necessity.

The 'frog' case is also settled by the 'safe indication' account. The method John is using does not safely indicate the beliefs truth, since John's method of inquiry could indicate anything irrespective of the truth of the proposition indicated. The 'safe basis' account

on the other hand disappoints us again, since the deliverance on which John bases his belief, is such that he would not have it unless his belief was true. Recall that John forms true beliefs in all the nearby possible worlds where he continues to believe that the frog is green. It should not matter that his belief is false in some faraway possible worlds where the evolutionary history of frogs has been different. The reason why John seems to lack knowledge in the case is because his method of belief-formation is ill-suited to form beliefs about the color of objects. This is something that the 'safe basis' account is unable to capture.

The 'brain lesion' case is a bit trickier. The trickiness arises from the fact that it is uncertain what the method of inquiry is. If we define methods externally, then we have to arrange the space of possible worlds so, that Alvin has the brain lesion in all nearby possible worlds. If this is the case, then it would seem that Alvin's belief satisfies 'safe indication' account, and Luper would be forced to grant Alvin with knowledge contra our intuitions. Alvin's method of belief-formation would not indicate that Alvin suffers from a brain lesion if he did not, mainly because the brain lesion is a vital part of the method. Be that as it may, we could formulate the methods internally, which would mean that we could rearrange the space of the possible worlds in such a way, that Alvin's belief no longer would satisfy 'safe indication'. But if the methods are to be defined on purely internal criteria, then the difference between the 'safe indication' and the 'safe basis' account, seems to vanish. For consistency's sake we should formulate the methods externally at least in the case of the 'safe indication' account. That being so the 'safe indication' account is in trouble with the 'brain lesion' case.

The 'safe basis' account on the other hand is much better suited to deal with the 'brain lesion' case. After all, Alvin could very well have the deliverance of suffering from a brain lesion, without it indicating the truth. If he suffered from severe headaches, he could very well believe that those headaches were caused by a brain lesion. So at least with respect to this case, the 'safe basis' account is superior to the 'safe indication' account.

But what about the 'red barn' case? Unfortunately the proposed conditions cannot handle the case. If Henry happens to be on a walk in the red-barn-façade-county and sees the red barn, his true belief that, there is a red barn in front of him, has a basis which it would not have if the belief was not true. Therefore Henry's belief is safe. Sosa

accepts this consequence: “the belief that one sees a real barn would seem safe after all, in the Kripke example, when deduced from one’s safe belief that one sees a red barn” (2004, 292).

The same applies to the ‘safe indication’ account. Henry’s method of belief-formation is deduction from his safe belief that he sees a red barn. The proposed conditions do not violate closure, since according to them one can both know that one sees a red barn and that one sees a barn. The problem with this kind of response is that it is highly counterintuitive that one could even know that one sees a red barn in the first place!

I am not sure whether Luper is bothered by this, but Sosa surely is not since he doesn’t advance ‘safe basis’ as a sufficient condition for knowledge, but merely as a necessary condition. According to Sosa:

> A belief must be not only safe but also virtuous if it is to amount to knowledge, where the virtue of the belief derives largely from the reliability of its sources. The fact that the belief about barns is safe does not automatically make it a case of knowledge, therefore, so it has not been shown after all that safety produces a counterintuitive result. (Sosa 2004, 293)

Even if this move allows Sosa to take a breath it will not save him. The reason for this is that intuitively, any coherent virtue-theoretic condition imaginable is already satisfied by the subject of the ‘red barn’ case. The reason for this is that the agent’s cognitive virtues are functioning perfectly, and yield a true belief. Nothing is amiss in the case from the perspective of the agent’s cognitive virtues.

8.3.2. Anti-luck virtue epistemology

Duncan Pritchard has advocated anti-luck virtue epistemology which exhibits ‘safety’ as a necessary condition for knowledge. It is a theory built around two ‘master intuitions’. According to the ‘ability’ intuition knowledge is always the product of one’s cognitive abilities while according to the ‘anti-luck’ intuition knowledge is incompatible with luck. Together these intuitions place two distinct demands on any normative theory of knowledge in such a way that they cannot be exhausted by a single condition. As a result we get a two folded theory of knowledge with an anti-luck condition and an ability condition. (Pritchard 2010, 51-52)

Pritchard cashes out the anti-luck intuition by the ‘safety’ condition and the ability intuition by a virtue theoretic condition:

Ability:
S’s true belief that *p* satisfies ‘ability’ if and only if:

(1) the belief arises out of S’s reliable cognitive traits that make up her cognitive character in such a way that S’s reliable cognitive success is to a significant degree creditable to her cognitive character. (Pritchard 2010, 54)

According to anti-luck virtue epistemology (ALVE) a subject S knows that *p* if her belief satisfies both ‘safety’ and ‘ability’. It should be easy to see that ALVE can deal with all the cases that ‘safety’ could deal with on its own. Notice also that there are some cases in which ‘safety’ is not satisfied, but ‘ability’ is. The barn façade cases, and thus cases featuring environmental luck, cannot be dealt by ‘ability’.

To see this consider the original ‘barn façade’ case . Henry’s belief that he sees a barn in front of him is not safe. There are many nearby possible worlds where he believes that he sees a barn, while his belief is false. Nonetheless his belief arises out of his reliable cognitive traits, namely from his perceptual skills. His perception is working normally and it gets things right. Pritchard agrees:

> [He] does not make any cognitive error in forming his belief in the way that he does. Accordingly, we would naturally say that [Henry’s] cognitive success is because of his cognitive ability, and so we would, therefore, attribute a cognitive achievement to [Henry]. That is, his cognitive success in this case is primarily creditable to his cognitive abilities. (Pritchard 2010, 35-36).

Virtue theoretic conditions like ‘ability’ are often ill-suited to deal with cases of knowledge where the subject of knowledge relies on other peoples cognitive skills or knowledge. Cases featuring testimonial knowledge are paradigm examples. Here is one introduced by Lackey in order to undermine virtue-epistemology:

SEARS TOWER:

> Having just arrived at the train station in Chicago, Morris wishes to obtain

> directions to the Sears Tower. He looks around, approaches the first adult passerby that he sees, and asks how to get to his desired destination. The passerby, who happens to be a Chicago resident who knows the city extraordinarily well, provides Morris with impeccable directions to the Sears Tower by telling him that it is located two blocks east of the train station. Morris unhesitatingly forms the corresponding true belief. (Lackey 2007, 352)

The point that Lackey is trying to establish with this thought experiment is that one can have knowledge without deserving credit for one's true belief. Indeed it seems obvious that Morris gains a piece of knowledge in the above case (notice that his belief is safe). Furthermore it seems obvious that his true belief is not primarily creditable to him. The credit falls to the helpful passer-by who knows the city extraordinarily well.

But unlike most virtue theoretic conditions, 'ability', does not demand that a belief that amounts to knowledge must be *primarily* creditable to the agent's cognitive character. Rather it demands that it must be to a significant degree creditable to her cognitive agency. But if this is the case, then it would seem that Morris's belief satisfies 'ability' after all. If we are to maintain the intuition that Morris has acquired knowledge in the case, then it seems plausible to think that he exercises at least some cognitive abilities that guide her belief formation. Morris would not have asked the directions from just about anyone, say a child or a vagabond, nor would she have been fooled by obvious lies. Clearly Morris must exhibit some cognitive abilities, and though he does not deserve primary credit for his true belief, he receives some, and that should be enough. (Prichard 2010, 41)

After briefly clarifying 'ability' we can move forward to pitch ALVE against the counter examples that 'safety' could not deal with in the previous section. We shall first look at the 'frog' case.

'Safety' gave the wrong verdict regarding the 'frog' case. There was no nearby possible world where John's belief that the frog is green was false simply because there were no non-green frogs in the environment in which John was. The question is whether 'ability' is satisfied in this case? It should be easy enough to see that John's belief is not in the least creditable to him, nor is it produced by a cognitive virtue that is suited for the task of forming beliefs about the color of objects. Simply put John does not have a reliable

cognitive trait to distinguish between colors, and therefore his belief cannot satisfy 'ability'. ALVE is apt to deal with cases like 'frog' where the direction of fit between the truth of the belief and the formation of belief is incidental.

What about the other cases? ALVE gives the right verdict regarding the 'brain lesion' case, at least if you are inclined to think that Alvin's true belief does not amount to knowledge. Although Alvin uses a reliable cognitive method to form his belief, it does not satisfy 'ability'. The reason for this is that the brain lesion is not a part of Alvin's cognitive character, nor is the skill it grants a trait. The brain lesion is actually a malfunction in Alvin's cognitive system, which just happens to generate true beliefs. It could have generated a myriad of false beliefs just as well. And because the brain lesion does not constitute a skill that is part of Alvin's cognitive character, a belief generated by it cannot satisfy 'ability'.

What about Kripke's 'red barn' case? Unfortunately ALVE cannot deal with this case any better than 'safety' could. This is because 'ability' is unable to deal with cases featuring environmental luck. If Henry happens to be in a red-barn-façade county, and forms the true belief that there is a red barn in front of him, his belief will satisfy 'ability'. The belief arises from his reliable cognitive traits, namely his perception, in such a way that the truth of the belief is primarily creditable to his cognitive skills.

So 'safety' and ALVE have a genuine counter example. Pritchard tried to evade the problems posed in the previous section by adding an ability condition that would ensure that the true belief rose from the agent's cognitive character. Now it should be evident that this strategy was doomed to fail. The reason for this is that the modal anti-luck condition should be able to deal with cases where an agent falls short of knowledge because of luck. The 'red barn' case is a classic example of environmental luck. 'Safety' should be able to deal with the case but it is unable. Adding a virtue theoretic condition will not help, since the subject lacks knowledge in the case because of luck, not because she has failed to exhibit cognitive virtues. This diagnosis is further strengthened as we analyze how ALVE deals with the 'malfunctioning calculator' case.

Actually the 'malfunctioning calculator' case is structurally similar with the 'red barn' case. To see this consider a scenario in which is the calculator is functioning perfectly. It seems obvious that I can gain knowledge with an ordinary calculator. Now how does ALVE deal with this case? 'Safety' is still automatically satisfied (which is of course a

bad thing) but ‘ability’ is also satisfied. Therefore it gives the right verdict and grants me with knowledge. In this mundane case my true belief is creditable to me, but most of the credit goes to the calculator.

Now consider the ‘malfunctioning calculator’ case. Intuitively my belief does not amount to knowledge. ‘Safety’ is still satisfied, but is ‘ability’ satisfied as well? Pritchard argues that it is not. Here is what he says on the issue:

> The problem posed by the [malfunctioning calculator] case is basically a restricted version of the more general problem that faces robust anti-luck epistemology, in that one can have a belief that exhibits all the right modal properties to satisfy the anti-luck condition (i.e. the safety principle, whichever way it is formulated) and yet which does not count as knowledge because of how the belief so formed is not in any significant way related to the agent’s cognitive ability. Given that this is the problem posed by the case, however, then it ought to be clear that, by incorporating a properly formulated ability condition, anti-luck virtue epistemology can deal with such cases. …It is clearly not the case that [the subjects] cognitive success is in any substantial way creditable to her cognitive character. (Pritchard 2009, 78)

I think that this is inconsistent. In a mundane case, where I use a perfectly well functioning calculator, I use exactly the same cognitive abilities that I use in the ‘malfunctioning calculator’ case. I file in the calculation correctly, I would have detected if the calculator was obviously broken and I would not have believed in just any answer. I use the same cognitive traits in both cases, and with as much success. Because I rely on an external device for my knowledge some or maybe most of the credit for my true belief fall to the calculator. However, some of the credit falls to me as well, just as Morris deserves some credit for his belief in the ‘Sears Tower’ case. The ‘ability’ condition is satisfied or dissatisfied in both the case where I use a calculator which is in working-order and in the ‘malfunctioning calculator’ case. I guess that Pritchard would have to agree with me that the condition is satisfied in both, because he has claimed that it is satisfied in the analogues ‘Sears Tower’ case (Pritchard 2010, 55). Furthermore Pritchard cannot claim that the ‘ability’ condition is not satisfied because the calculator is not a reliable device by which to reach true belief in the case described.

This is because as an external device the malfunctioning calculator is not a part of the agent's cognitive character in the first place. Once this is realized it is even more obscure how a condition like 'ability' could save ALVE from counterexamples like 'malfunctioning calculator'.

Since both 'safety' and 'ability' are satisfied in the 'malfunctioning calculator' case, I have knowledge according to ALVE in the case. This is counterintuitive and therefore ALVE has a second counterexample. Despite what Pritchard's claims, ALVE is incapable of dealing with cases featuring necessary truths. Adding a virtue theoretic condition to 'safety' will not mend the problem. We will have to look at radically new ways of formulating our modal condition for knowledge.

9. Global safety

At this point it should be evident that 'safety' as well as 'sensitivity' faced a fair deal of problems. Still I think that it is safe to say that 'safety' fared better in comparison. Yet 'safety' does not offer us what we were looking for. There might be a better analysis of knowledge on our very doorstep. After critically evaluating different proposals at length I will now turn to argue for my own position.

All the cases that were problematic for 'safety' shared a common element. In the 'brain lesion' case, 'red barn' case, 'frog' case and the 'malfunctioning calculator' case the belief formed by the subject in the actual world either couldn't have been wrong or would not easily have been wrong. Since 'safety' demanded that a belief qualifies as knowledge, if and only if it couldn't have easily been wrong, safety was satisfied. Nevertheless the subjects in the above mentioned cases could very easily have erred. Their specific beliefs might have been safe from error, but they themselves as epistemic agents, were not safe from error.

It is generally thought that knowledge either demands certainty of a certain level, absolute certainty, or a capacity to acquire infallibly true beliefs. Though I am not advocating infallibilism, I am quite ready to accept the intuitive force behind their view. In fact the whole project of offering a modal condition for knowledge can be seen as an attempt to formulate a condition which requires that a subject's beliefs must satisfy a certain level of certainty in order to aspire to knowledge. To achieve this, it is not

enough that one's belief is safe from error, oneself must also be safe from error. 'Safety' was formulated so that it concerned only whether the belief formed in the actual world was true. This of course will not ensure that one is safe from error, only that one's particular belief that *p* is safe from error. We must reformulate 'safety' in such a way that it ensures that the subject is safe from error. To do this, the formulation will need to take into account not only a specific belief, but a set of possible beliefs that the agent forms in nearby possible worlds. In order for a subject to have knowledge the relevant set of beliefs formed by the agent in nearby possible worlds has to contain only true beliefs (or a certain proportion of the beliefs have to be true).

Global safety: *S* knows that *p* (which belongs to the set of propositions *P*) if and only if

(1) *p* is true, and *S* would not have easily formed a belief in any given proposition that is false which belongs to *P* (by the same method that she uses in the actual world to form her belief).[16]

This should be read as follows in terms of possible worlds:

Global safety*: S knows the true proposition that *p* (which belongs to the set of propositions *P*) if and only if

(1) in nearly all nearby possible worlds where S believes in a proposition belonging to *P* (by the same method M that she uses in the actual world), S's belief is true,
(2) and in all of the very closest nearby possible worlds where S believes in a proposition belonging to *P* (by the same method that she uses in the actual world), S's belief is true.

This formulation captures nicely the intuitive idea that was behind 'safety' and formulates it correctly unlike its predecessor. It is not enough that a particular belief of ours is safe, our epistemic status has to be such, that we could not have erred with respect to other, similar beliefs, in which we would have believed in nearby possible worlds. If a subject satisfies 'global safety' then she is, as an epistemic agent, safe from error regarding the subject matter of her belief.

[16] All the propositions that belong to the set of propositions *P* have to be related in some sense. They must be about the same subject matter. For example, in the calculator cases they would have to be mathematical propositions, and in the barn façade cases they would have to be about barns. We do not want to claim that our subject does not know that the sun is shining simply because he is in a barn façade county!

9.1. Four fatal problems revisited

How does ‘global safety’ handle the four fatal problems? It should be quite clear that it can deal with all of them. Consider the ‘malfunctioning calculator’ case. In this case the subject forms a belief in a necessarily true proposition by using a malfunctioning calculator. The subject’s belief is true, but there is a number of nearby possible worlds, where she forms a belief, with the same method she uses in the actual world, in a false proposition, that is about the same subject matter as her actual belief. Thus she does not satisfy ‘global safety’ and does not possess knowledge as our intuitions dictate.

The ‘frog’ case is also easily dealt. Even though John does not falsely believe that the frog is green in any nearby possible world (simply because all the frogs in his environment are green) he will form a variety of false beliefs regarding the frogs color in nearby possible worlds. Thus John lacks knowledge.

The ‘red barn’ case is dealt with similar elegance. Because the subject of the ‘red barn’ case will form a plentitude of false beliefs regarding barns in nearby possible worlds, she does not know that there is a red barn in front of her. We noted earlier that this result is the most preferable one, since it is the intuitively most correct answer to the thought experiment and it allows us to preserve closure. Thus ‘global safety’ fares much better regarding the ‘red barn’ case than for example Sosa’s ‘safe basis’ account, which allowed us to maintain closure but granted knowledge to the subject of the case.

The ‘brain lesion’ case is a bit trickier. First of all it is not altogether clear that Alvin fails to know in the case that he suffers from a brain lesion. Consider the positive factors regarding Alvin’s belief. First of all the belief is produced in a reliable way. Secondly it satisfies both ‘safety’ and ‘sensitivity’. It seems to satisfy the modal constrains that knowledge requires. What is wrong with the belief? Apparently its source, and the fact that Alvin does not know or understand why he believes as he believes. This would indicate that an internalistic condition for knowledge is needed. Perhaps that is so. Or might it be that Alvin satisfies a kind of brute externalistic conception of knowledge, and that the internalistic demands are required for a higher epistemic standing, say understanding or reflective knowledge? Sosa has divided the concept of knowledge into animal and reflective knowledge. Here is a brief characterization given by Sosa of the notions:

> Human knowledge has at least two varieties, the animal and the reflective: 'knowledge' sometimes means the first sometimes the second. This is not necessarily to say that the word itself is ambiguous in English… In any case, animal knowledge does not require that the knower have an epistemic perspective on his belief, a perspective from which he endorses the source of that belief, from which he can see that source as reliably truth conducive. Reflective knowledge does by contrast require such a perspective. (Sosa 2011, 135)

In Sosa's terms our project should then be seen as finding the conditions for animal knowledge, which does not require a perspective into ones cognitive character. Therefore it should not be all too problematic to grant Alvin with knowledge in the case since we are only granting him with animal knowledge.

Furthermore, depending on how the details of the case are put together, a proponent of 'global safety' can claim that Alvin does not know that he has a brain lesion. This is because there might be nearby possible worlds where the brain lesion produces other, false beliefs. Therefore, as I see it, a proponent of 'global safety' can deal with the 'brain lesion' case in either way. Ironically this can be seen as a virtue of the view, since intuitions about this case are shifty. To determine whether 'global safety' is satisfied by Alvin or not, we would have to know more about the nature of the brain lesion.

'Global safety' fares exceptionally well with these four cases. The responses it gives to the other thought experiments that have been introduced earlier are also satisfactory and very much the same that 'safety' gave. The reader is free to test the mettle of 'global safety' against those cases. Now we will turn to look at two proposals that are somewhat similar to 'global safety'.

9.2. Similar proposals

As it happens 'global safety' is not all that novel condition for knowledge. We will briefly examine Pritchard's enhanced ALVE and an alternative way of improving the 'safety' condition, namely marrying it with Nozick's 'adherence' condition. All of these

proposals have something in common with ‘global safety’ though they are not equivalent with it.

9.2.1. Enhanced ALVE

As it happens ‘global safety’ is not all that novel condition for knowledge. Pritchard has suggested that one could refine the ‘safety’ condition by focusing on the doxastic output of the belief-forming method. “The issue, then, would not be whether there is a nearby possible world where one believes the target proposition on the same basis as in the actual world but where one’s belief is false, but rather whether there is a nearby possible world in which one employs the same belief-forming method as in the actual world and thereby comes to form a false belief as a result.” (Pritchard 2010, 57-58)

Of course Pritchard’s refinement of ‘safety’ would need even more refinement in order not to fall prey to obvious counter examples. He would have to, as we have done, restrict the domain of beliefs that count as relevant when determining whether his condition is met. Otherwise he will end up claiming that any belief produced by a belief-formation method, will not count as knowledge, unless the belief-formation method is reliable regardless of the subject matter. For example, if I were in a barn façade county, and looked at the sun and formed a true belief that the sun is shining, then I ought to know that the sun is shining. Nevertheless I could have easily formed a multitude of false beliefs (about barns) using the very same belief-formation method, namely perception, but this fact should not rob me of knowledge. By restricting the condition to beliefs that are about the same subject matter as the belief believed in the actual world this problem is sidestepped.

But Pritchard does not see this refinement as saving ‘safety’ as a necessary and sufficient condition for knowledge, or even as a required modification to his ALVE account of knowledge. The refinement only postpones the problem that ‘safety’ is about to face, since one can always restate the same problem so that the agent will not form any false beliefs in nearby possible worlds because, say a wizard ensures that the subject’s beliefs are always true (Pritchard 2010, 58). Pritchard claims that by incorporating a virtue theoretic condition, like ‘ability’ one can evade these problems, but it should be clear by now that this move does not succeed. Indeed it seems that the

‘ability’ condition is largely redundant if we accept a condition like ‘global safety’. ‘Global safety’ does not need to be enhanced by a virtue theoretic condition.

Thought experiments that feature wizards, helpful demons and guardian angels should not in my mind be taken too seriously. Therefore I do not consider such examples as potential counter examples to epistemological theories. This move is not *ad hoc*, and I will give grounds for it later when we ponder the role that the concept of knowledge plays in our society. For the moment however the reader must contend with the brute response to bizarre thought experiments.

9.2.2. ‘Safety’ and ‘adherence’

Recall the ‘adherence’ condition ? By allying ‘adherence’ with ‘sensitivity’ the tracking theorist could evade the problems that necessary truths posed. Since the ‘safety’ principle also had trouble with necessary truths and analogues cases, it would be natural to suppose that by demanding ‘safety’ as well as ‘adherence’ one could save ‘safety’.

Alas, even though ‘safety’ and ‘adherence’ together can deal with all the thought experiments introduced so far, it is clear that ‘adherence’ is not adequate to the task at hand. The reason for this is, that adherence focuses only on the subject’s actual belief, and demands that the subject holds that belief throughout nearby worlds where it continues to be true. This is not enough, since knowledge seems to require that we are in a state where we could not easily have erred, not only about the particular belief in which we actually believed, but about all the beliefs in which we believe in nearby possible worlds that are about the same subject matter. Only a condition like ‘global safety’ can cash this out. To demonstrate how ‘adherence’ plus ‘safety’ fail consider a variation to the ‘malfunctioning calculator’ case.

TWISTED CALCULATOR:

In this case I reach the true belief that 12 x 13 =156 by using a malfunctioning calculator. However, the malfunction is such that the particular calculation I made is unaffected by the malfunction. Had I entered any other calculation, the output would not have been true. So with respect to the calculation I made, the machine was working properly.

Intuitively this is not a case of knowledge. The calculator is broken. By a happenstance the particular calculation I made is unaffected by the malfunction. Therefore in all the nearby possible worlds where I try to calculate the product of 12 x 13, I form a true belief. 'Adherence' and 'safety' are met in this case, though 'global safety' is not met. Therefore only 'global safety' characterizes correctly the way in which a subject must be safe from error in order to attain knowledge.

Now we will turn to consider the still threatening counter examples and other perhaps even more severe and fundamental problems that the advocate of 'global safety' faces.

9.3. Hit and run counter examples

After 'safety' gained popularity as a necessary condition for knowledge, counter examples were introduced, that were specifically tailored to show that any condition like 'safety' cannot constitute a necessary condition for knowledge. Because 'global safety' shares a lot with 'safety' it too is susceptible to these counter examples. We will now turn to evaluate whether they indeed threaten our analysis of knowledge, or can they be evaded.

Steven Luper has noted that some of the cases that threaten 'safety' feature 'restorative rigging'. In standard Gettier cases, a subject uses a method, which in normal circumstances reliably indicates the truth of her belief, yet the method happens to be unreliable in the subject's actual environment for some peculiar reason, to arrive at a true belief. 'Safety', and thus 'global safety', were apt to deal with this kind of 'deleterious rigged' scenarios. The cases that threaten 'safety' are however restoratively rigged; the reliability of the method used by the subject is restored by some other happenstance, making the method reliable again. (Luper 2006, 161)

Here is a case involving restorative rigging:

PAPIER-MÂCHÉ BARNS

> An artist is about to implement a plan to spread papier-mâché barns throughout the neighborhood in which [Steven] will later walk. The fakes are loaded in a truck, ready to go. But the artist has several friends who think such projects are absurd, and convince him to change his mind. So

> no fakes are in the neighborhood, It is still later that [Steven] enters the area, stands before a barn, and forms a visual impression of it in optimal viewing conditions, coming thereby to believe a barn is there. (Luper 2006, 164)

This case seems to be a clear case of knowledge, although some philosophers, such as Shope (2002), think that cases featuring restorative rigging can never constitute knowledge. Restorative rigging gives rise to diverging intuitions in epistemologists, but as we will later see, this works in our advantage. Be that as it may, 'global safety' would indicate that Steven knows that there is a barn in front of him. There are no nearby possible worlds where he forms false beliefs about barns, since the artist's plans have changed. "Figuratively speaking, the deleterious rigging […] 'pulls' worlds in which S is deceived closer […] while the restorative rigging 'pushes' them back out" (Luper 2006, 164). In this case it seems that the restorative rigging 'pushes' the worlds that would bar Steven from knowing far enough. In other cases our intuitions might be more conflicted. At this point it is important to note that 'global safety' can handle cases featuring restorative rigging in either way, and because restoratively rigged cases are often borderline cases of knowledge, this will play in the hands of the advocates of 'global safety'.

Here is a case offered by Neta and Rohrbaugh which is supposed to show that a condition like 'safety' cannot be a necessary condition for knowledge.

WATER

> I am drinking a glass of water which I have just poured from the bottle. Standing next to me is a happy person who has just won the lottery. Had this person lost the lottery, she would have maliciously polluted my water with a tasteless, odorless, colorless toxin. But since she won the lottery, she does no such thing. Nonetheless, she almost lost the lottery. Now, I drink the pure, unadulterated water and judge, truly and knowingly, that I am drinking pure, unadulterated water. But the toxin would not have flavored the water, and so had the toxin gone in, I would still have believed falsely that I was drinking pure, unadulterated water. (Neta & Rohrbaugh 2004, 399)

Neta and Rohrbaugh think that this is a clear case of knowledge. Luper (2006, 167) sees it as a borderline case of knowledge, while Pritchard (2012a, 185) does not even consider it as knowledge. Where intuitions diverge we can only hope that 'global safety' allows us to deal with the case either way, depending on how the details are put together. Luckily this is the case.

While it is not clear whether there is not knowledge in 'Water' it is also not clear that 'global safety' is not met as Neta and Rohrbaugh would have us believe. The reason for this is that it is unclear how far the deceptive worlds are pushed back by the restorative rigging. Neither is it clear how far is enough. Since these are borderline cases of knowledge this is totally acceptable.

Let us look at another case that might invoke clearer intuitions.

HALLOWEEN PARTY

> There is a Halloween party at Andy's house, and I am invited. Andy's house is very difficult to find, so he hires Judy to stand at a crossroads and direct people towards the house (Judy's job is to tell people that the party is at the house down the left road). Unbeknownst to me, Andy doesn't want Michael to go to the party, so he also tells Judy that if she sees Michael she should tell him the same thing she tells everybody else (that the party is a the house down the left road), but she should immediately phone Andy so that the party can be moved to Adam's house, which is down the right road. I seriously consider disguising myself as Michael, but at the last moment I don't. When I get to the crossroads, I ask Judy where the party is, and she tells me that it is down the left road. (Comesaña 2005, 397)

Comesaña claims that he ought to know where the party is, and furthermore that his belief does not satisfy 'safety' or 'global safety'. According to Comesaña there are nearby possible worlds where he forms a false belief, while believing that the party is down the left road.

I agree that there is knowledge in the case. However it is far from obvious that the advocates of a safety condition cannot deal with this case accordingly. In fact there are

several ways to elude the problem posed by this case.[17] First of all notice that the case is restoratively rigged. Therefore, depending on how far the restorative rigging pushes the deceptive worlds, the stronger our intuition of knowledge in this case is. After deciding not to dress up as Michael I would be inclined to say that the deceptive worlds are pushed very far away indeed. The proponents of 'safety' or 'global safety' can therefore endorse the intuitive response to this case.

As if this was not enough it should be easy to see that the case never actually threatened any kind of safety account. To see this, think about the belief formed by the subject of the case. Whether he is dressed up as Michael or not he will form the belief that the party is down the left road. As it happens, this belief is true whether or not the subject is dressed up as Michael or not. If he is dressed up as Michael, then the truth of his belief will change in a short while, but that does not undermine 'safety'. If I look out of the window and see a bird there, and my belief is safe, I know that there is a bird, even if the bird soon flies off. If I continue to believe that there is a bird outside my window (while it has flown off) without seeing the bird, my belief is simply false. The truth value of the belief changed as time flew forward, but this does not affect whether my original belief was safe or not, nor whether it was knowledge or not.[18]

Here is the final thought experiment that is supposed to show that 'global safety' is not a necessary condition for knowledge.

ATOMIC CLOCK

> [T]he world's most accurate clock hangs in Smith's office at a cereal factory, and Smith knows this. The clock's accuracy is due to a clever radiation sensor, which keeps time by detecting the transition between two energy levels in cesium-133 atoms. This radiation sensor is very sensitive, however, and could easily malfunction if a radioactive isotope were to decay in the vicinity (a very unlikely event, given that Smith works in a cereal factory). This morning, against the odds, someone did in fact leave a small amount of a radioactive isotope near the world's most accurate

[17] Tomas Bogardus argues that the case does not pose a problem for the advocates of 'safety' on different grounds then we do below. (Bogardus 2012, 10)

[18] According to a standard interpretation of propositions their truth values do not in fact change, but sentences uttered at different times (and locations) express different propositions. Similarly sentences believed in at different spatiotemporal locations express different propositions. This fact bears no consequences on the above argument.

> clock in Smith's office. This alien isotope has a relatively short half-life, but—quite improbably—it has not yet decayed at all. It is 8:20 am. The alien isotope will decay at any moment, but it is indeterminate when exactly it will decay. Whenever it does, it will disrupt the clock's sensor, and freeze the clock on the reading ''8:22.'' (Don't ask why; it's complicated.) Therefore, though it is currently functioning properly, the clock's sensor is not safe. The clock is in danger of stopping at any moment, even while it currently continues to be the world's most accurate clock. (Bogardus 2012, 12)

The question is does Smith know that it is 8:22 when he comes to his office? Bogardus claims that he knows, and that 'safety', whichever way it is formulated, is not fulfilled. Therefore 'atomic clock' presents a genuine case of unsafe knowledge, at least according to Bogardus. Bogardus backs this claim by highlighting the virtues of Smith's belief. Smith's belief is reliably produced, the belief manifests from his intellectual virtues, the belief is appropriately connected (in a causal manner) to its truth, etc. (2012, 13). Clearly there is much good to say about Smith's belief. Nonetheless all of the theories that Bogardus quotes as granting Smith with knowledge have well known counter examples.

Moreover the intuition that Smith knows that the clock is 8:22 is far from unanimous. Pritchard for example, discussing an almost identical case, does not share the intuition at all. According to him Smith is essentially finding the time by looking at a stopped clock (Pritchard 2012a, 187). I am inclined to agree with Pritchard on this score. I find it difficult to accept that the clock is 'the world's most accurate clock' since it will stop at any moment. Clearly the clock would be more accurate if it was not going to stop in the matter of minutes or seconds!

However it would be nice if we could accommodate Bogardus as well as Pritchard's intuitions. Luckily we can. 'Atomic clock' is also a case that has been restoratively rigged, but in a quite peculiar manner. The radioactive isotope in the room deleteriously rigs the case. Quite improbably the isotope has not decayed yet at all. The fact that the isotope has not decayed when Smith looks at the clock restores the clock's reliability, if only temporarily. If we were not told that isotope had not decayed we would have judged that Smith does not know that the clock is 8:22. But since we are granted this

information, we can see that the deceptive worlds are pushed far enough to perhaps grant Smith with knowledge.

The more complicated the cases get, the more our intuitions seem to diverge when determining whether the subject of the case has knowledge or not. It should be seen as a virtue of 'global safety' that it can explain these diverging intuitions. Now we will turn to look at some of the consequences of accepting 'global safety' as a necessary and sufficient condition for knowledge would have.

9.4. Skepticism rebutted

One of our key goals was to produce a theory of knowledge that could affirm most of our everyday knowledge attributions as true. By being able to handle all the thought experiments that have been presented so far, 'global safety' seems to be able to fulfill this wish. But we have only hinted at the anti-skeptical consequences of 'global safety'. Now we will focus on them.

The greatest difference between 'sensitivity' and 'safety' was that 'sensitivity' entailed that we do not know the denials of skeptical hypotheses. 'Safety' and 'global safety' on the other hand entail that we do know them. This is because there are no nearby possible worlds where we are radically deceived. The faraway possible worlds have no bearing whatsoever on whether we have knowledge or not. The anti-skeptical consequences of 'global safety' are great indeed. Skepticism is rebutted and should not concern us anymore.

But this is not right! Knowing that we are not deceived by a malicious demon or that we are not just hallucinating should not be so easy! According to 'global safety' it is simply enough to believe that one is not the victim of demon, in order to know that one is not. Acquiring such knowledge should require sophisticated *a priori* arguments! While the anti-skeptical consequences are welcome, the advocates of 'global safety' have to explain how this sort of metaphysical knowledge is so easy to obtain. This is the so called 'problem of easy knowledge'. Is 'global safety' satisfied all too easily in the case of skeptical scenarios?

I am going to take a neo-Moorean stance to skepticism. Sosa, Pritchard and Greco have all defended a neo-Moorean answer to skeptical puzzles to name a few contemporaries

(Sosa 1999,147-148), (Pritchard 2002a, 2002b, 2012b, part three), (Greco 2010, 174-196). To better illuminate the nature of the neo-Moorean stance we should consider the neo-Moorean response to the following skeptical argument:

Radical skepticism

(1) I do not know that I am not a BIV.
(2) If I know that I have two hands, then I know that I am not a BIV.
(3) I do not know that I have hands.

The motivation for (1) comes from reflecting on the nature of radical skeptical scenarios and the motivation for (2) comes from the highly intuitive closure principle (Pritchard 2012b, 112). (3) follows by deduction from (1) and (2), so the argument is formally valid. The neo-Mooreean's response to the above argument is to deny the first premise. More specifically the neo-Moorean just turns the skeptic's *modus tollens* argument into a *modus ponens* argument.

Neo-Mooreanism

(1) I know that I have two hands.
(2) If I know that I have two hands, then I know that I am not a BIV.
(3) I know that I am not a BIV.

In addition to providing this ingenious argument, the neo-Moorean also gives some theoretical motivation for his argument. 'Global safety' provides us with theoretical motivations for accepting the neo-Moorean argument. After all, from 'global safety' it follows that I know that I have hands, and by closure I can reason from there that I am not a BIV. However since the skeptical hypotheses are paradoxes, with plausible premises that taken together yield an absurd conclusion, we have to explain why the false premise seemed so plausible. "We must explain how two premises that together yield a conclusion we find so incredible can themselves seems so plausible to us. Only with such an explanation in place can we proceed with confidence and with understanding to free ourselves from the [skeptic's] trap" (DeRose 1995, 3).

So our mission is quite clear now; we have to figure out why the skeptical hypothesis's first premise seems so plausible. Luckily there are several different strategies available

to us and they are not mutually incompatible. Perhaps we find the skeptic's first premise plausible for several reasons.

Sosa's explanation relies on the fact that 'safety' and 'sensitivity' are easily confused with each other, since they are mutual contrapositives (1999, 148). Recall that 'sensitivity' in its plainest form could be expressed as $\sim p \rightarrow \sim B(p)$ and 'safety' as $B(p) \rightarrow p$. While 'global safety' is not equivalent to the brute formulation of 'safety' it is still very similar to it. 'Global safety' gives the necessary and sufficient conditions for knowledge, while 'sensitivity' is not required of knowledge. Furthermore 'sensitivity' would deny that I know that I am not a BIV, since in the closest possible worlds where I am, I still believe that I am not. If 'sensitivity' and 'global safety' are easily confused with each other, then it should be easy to see why many find the skeptic's first premise so plausible. They simply confuse 'sensitivity' with the correct condition for knowledge, which is given by 'global safety'.

Sosa's explanation is quite satisfactory, but there are several other ways to explain the plausibility of the skeptic's first premise. This only strengthens our neo-Moorean stance. Pritchard has offered another kind of answer as to why the skeptical paradoxes first premise seems so plausible.

Pritchard's strategy is to undermine the very thought that premise (1) ever was intuitive. He claims it to be intuitive only if you have accepted misguided internalistic philosophical theories about knowledge. Thus Pritchard's strategy is stronger than Sosa's was. Sosa's strategy was an over-riding one. Sosa gave an explanation as to why (1) is intuitively plausible and then proceeded to give independent theoretical motivations to abandon it. The over-riding strategy will leave the paradox intact (it will not render (1) unintuitive), though harmless, while an undercutting strategy will dissolve the whole paradox as mere fantasy (Pritchard 2012b, 133).

Pritchard explains our hesitation to utter claims that deny (1) by relying on conversational implicatures carried by our knowledge attributions. According to Pritchard, a perceptual knowledge claim that is made as a response to a specific

challenge, conversationally implicates, that the subject can discriminate between what she asserts and the error possibility raised (Pritchard 2012b, 142).[19]

The propositional content of an assertion can be true, even though the assertion implies something which is in fact false. If an explicit knowledge claim that is used as a response to a specific skeptical challenge (say the BIV hypothesis) conversationally implies that the utterer is able to distinguish between the actual world and the BIV world, then it implies something which is false. After all, by hypothesis we are unable to distinguish the BIV world from the actual world. This however does not affect the truth value of the original knowledge claim (I know that I am not a BIV), it only makes it conversationally inappropriate, because implying something that is false is considered quite rude. Pritchard can quite satisfactorily explain our reluctance to utter explicit knowledge claims that state that we know the denials of skeptical hypotheses.

Lastly I will only hint at a third way of solving the problem of easy knowledge. By endorsing contextualism one can evade the problem altogether. According to contextualism we can know the denials of skeptical hypotheses in quotidian contexts, but if we move to highly skeptical contexts we will not know anything, since the bar for knowledge rises incredibly high. The meaning of 'knows' shifts with the conversational context. As a matter of fact contextualism is particularly well fitted to 'global safety'. One can explain the shifts in contexts as shifting the boundaries of 'nearby possible worlds'. In normal contexts the nearby possible worlds that count as relevant when determining whether S knows that *p* are quite near to the actual world. In more skeptical contexts the boundaries are expanded to include even the worlds that are normally considered to be very far away. Thus contextualists can claim that we sometimes do not know the denials of skeptical hypotheses. Knowledge is sometimes easier to come by, sometimes next to impossible.[20]

It is fair to say that 'global safety' can handle the problem of easy knowledge. There are however some, more technical problems that still need solving.

(Maybe something about the nearness of possible worlds, or maybe it is enough to mention them in the reductive analysis part?)

[19] Although Pritchard focuses solely on perceptual knowledge, I do not see why the strategy he offers could not be used in a more general account of knowledge with some minor modifications.
[20] Keith DeRose has pursued this line of reasoning, the difference being that he advocates 'sensitivity' rather than 'safety' as a feature of knowledge (DeRose 1995).

10. Role of knowledge

At the beginning of our project we noted that a theory of knowledge is successful only if it can explain why knowledge has the suggested structure. Why is *this* concept so useful, that it is found universally in almost every society? Our focus will be on Edward Craig's genealogical analysis of knowledge and how it fits together with 'global safety' and whether it supports it or not.

According to Craig's genealogical analysis the concept of knowledge has evolved to flag good informants (1990, 11).

> Human beings need true beliefs about their environment, beliefs that can serve to guide their actions to a successful outcome. That being so, they need sources of information that will lead them to believe truths. They have 'on-board' sources, eyes and ears, powers of reasoning, which give them a primary stock of believes. It will be highly advantageous to them if they can also tap the primary stocks of their fellows - the tiger that Fred can see and I can't may be after me and not Fred… On any issue some informants will be better than others, more likely to supply a true belief… So any community may be presumed to have an interest in evaluating sources of information; and in connection with that interest certain concepts will be in use. (Craig 1990, 11)

This is the starting point of Craig's genealogical analysis of the concept of knowledge. His aim is to find a concept that allows human beings to distinguish other members of their community as good informants regarding a given subject matter. Take note that this strategy aims to explicate only a proto-concept of knowledge, since the concept searched by Craig is not yet suited to attribute knowledge to one self. The main quest is to find a detectable property of a potentially good informant that correlates well with having true beliefs about a certain subject matter.

At first we must note that in order to be a good informant, one has to form true beliefs about the matter at hand, not only in the actual world, but also in possible worlds. The reason for this is, that even though an inquirer is interested in what the truth of the

matter at hand is in the actual world, she does not know what kind of world the actual world is. From the perspective of the inquirer the actual world can be any of the possible worlds which she has not yet ruled out. Finding an informant who holds true beliefs in many counterfactual situations is therefore clearly advantageous. (Craig 1990, 19-20)

But how does the inquirer distinguish that a potential informant holds true beliefs about the matter at hand in most nearby possible worlds? Craig suggests the following line of thought:

> One possibility is that he [the inquirer] has found the potential informant to be right on this question in the past, and so believes that he will be right about it this time too. In this event it would be very unusual for his belief to be limited to just this one occasion; nearly always, if not actually always, he will believe that the potential informant is, and will be until some drastic change alters the situation, in general right as to whether *p*. But once he believes that, then he believes something about the informant's capacities beyond the present case… (1990, 56).

So the detectable property that correlates well with being right about a given subject matter is that the informant has been right about the same sort of issues in the past. This seems to be extremely plausible. If I want to know whether there is a tiger lurking nearby I will most probably consult Fred who is on the lookout and has reliably detected tigers before.

'Global safety' is particularly well suited to accommodate this characterization of a good informant. First of all an informant that satisfies 'global safety' regarding some subject matter will hold mostly true beliefs in nearby possible worlds regarding the subject matter. Secondly an agent that satisfies 'global safety' will most probably have had true beliefs about the matter at hand in the past also. After all, she forms only true beliefs about the matter in possible worlds. It is plausible that an agent who satisfies 'global safety' regarding a certain set of propositions will have satisfied it in the past also. To be in a globally safe epistemic status requires comprehensive skills and knowledge. Of course there are exceptions, but this should not put us of. According to Craig the way 'things nearly always go' enjoys a special status in the context of his

analysis, since the inquirer's context is a practical one (1990, 15).[21] What matters for us is that 'global safety' often goes hand-in-hand with a detectable property that correlates well with having true beliefs about a subject matter.

What is important to note here, is that we are talking about a subject matter, not about a particular proposition. Craig notes this too, and makes a concession to views that analyze knowledge not in terms of a single proposition, but in terms of a group of propositions (1990, 56). This grants 'global safety' and edge over the rival modal conditions, since it can account for the fact that an inquirer is often not interested only in whether *p*, but in a range of propositions. If I ask Fred what kind of weather it is outside, I will want him to be able to tell me not only whether it is raining, but also whether it is windy, snowing, foggy, sunny and so on. 'Global safety' is much better suited to explain this feature than 'sensitivity' or 'safety' was.

A proponent of 'global safety' can quite happily retort to the plausible idea that the concept of knowledge is primarily used to pick out good informants. Furthermore Craig's analysis allows us to give a good explanation as to why knowledge has the particular structure as offered by 'global safety'. Craig has given us independent grounds to think that the term 'knowledge' has evolved to pick out good informants. As it happens, informants that we would consider as good informant most often satisfy 'global safety'. Knowledge has the proposed structure because the concept of knowledge is used to flag good sources of information and good sources of information tend to satisfy 'global safety'.

Notice also that Craig's genealogical analysis only strengthens our position against the skeptic. Since the concept of knowledge has evolved to pick out good informants, informants who hold true beliefs in most nearby possible worlds, it would be odd that the skeptical scenarios, which are realized only in faraway possible worlds, would have any effect on the truth of our knowledge attributions. According to Craig's conception of knowledge, we do not have to worry about the possibility that an informant will form false beliefs in some possible worlds that are faraway form the actual world. 'Global

[21] Notice that this particular aspect supports our decision to abandon the extraordinary weird counterexamples. Since the context of acquiring knowledge is bound to be a practical one, we should not consider thought experiments which could not be realized in a world that is like ours. Wizards, demons and hidden helpers are all too farfetched to have any effect on our knowledge attributions. Therefore our intuitions regarding such cases should not be taken as truth entailing. Therefore the move to reject such thought experiments is not *ad hoc* at all.

safety' echoes this. We have found yet another reason to adopt the neo-Moorean stance on skepticism.

11. Dodged problems and topics for future research

At the beginning we noted that a plausible theory of knowledge will tell us why knowledge is more valuable than a belief that is merely true. As it turns out I am running out of space to address this important question properly. Here I will only hint at a solution to the value problem. According to my view, knowledge is often, but not always, more valuable than true beliefs that fall short of knowledge. I think it is quite plausible to think that knowledge is valuable because it is an achievement and achievements are generally valuable to those who achieve them. Nevertheless knowledge is not only a cognitive achievement, as the barn façade case illustrates. Recall that in that case a virtue theoretic condition was satisfied even though the subject did not have knowledge. I believe that cognitive achievements and 'global safety' go very often hand in hand. If you satisfy 'global safety' then you will most probably exhibit some sort of cognitive achievement also. This might not necessarily be the case, and that is partly why I don't think that knowledge is always of greater value than beliefs that fall short of it. Regrettably I don't have the space to dwell on the value problem here. Resolving the value problem of knowledge from the perspective of modal epistemology is definitely a topic for future research.

In addition to the value problem there are three issues that we have either dodged or postponed this far. These are:

(1) Where does the border of nearby and faraway possible worlds lie and how is it determined?
(2) How should the generality problem be solved, or can it be solved?
(3) How is the set of propositions 'P' determined and what propositions belong to the same 'subject matter'?

Problems (1) and (2) are of a more general sort, and many theories of knowledge struggle with them. Recall however that problem (1) is not all that fatal for us, since we

have also given a definition of ‘global safety’ without mentioning possible worlds. However it would be nice if we had an explicit way of demarcating worlds near and far. Problem (3) is of course specific to ‘global safety’, but I don’t think that it is all that pressing. What belongs to a certain subject matter is most often fixed by the context where knowledge is attributed or denied. Even though the problems above are severe I do not think they are fatal. They all seem to arise from vagueness. It is indeterminate how the border between worlds near and far should be drawn and the whole concept of possibility seems to be vague. We can live without precise answers to these questions; all we need is a picture, or a general understanding, of the possibilities expressed. In the same vein I think that the generality problem arises due to vagueness. If the vagueness of methods, subject matters and the metrics of possible worlds are in synch with the vagueness of the predicate ‘knows’ then we can be content with this response. Williamson echoes this point (2000, 100). However this quick response to these problems obviously falls short and will probably not satisfy many critics. Here I have only hinted at where the possible solution might reside. Examining how the vagueness of the predicate ‘knows’, correlates with other vague expressions that are critical to our definition of knowledge, would be a topic for future investigation.

Modal theories of knowledge are not often used in a social epistemology, which is usually considered to be the domain of evidentialism and dialectical approaches. It would however be extremely interesting to test the mettle of ‘global safety’ regarding the problem of peer disagreement. At a glance ‘global safety’ would seem to fare far better than its rivals (e.g. ‘sensitivity’ and ‘safety’) regarding that problem.

Another topic for future investigation would be to apply ‘global safety’ to the study of metaphysical knowledge of necessary truths. Since ‘global safety’ can aptly deal with necessary truths (i.e. a belief in a necessary truth does not automatically constitute knowledge according to ‘global safety’, contra ‘sensitivity’ and ‘safety’) the prospects of it regarding this problem are brighter than those of competing theories of knowledge. Clearly there are many ways in which ‘global safety’ could be put to good use.

12. Concluding Remarks

Thus we have arrived at a modal theory of knowledge that is able to deal with all of the presented counter examples in a quite satisfactory manner. More importantly, 'global safety' seems to cash out the anti-luck intuition. It seems to capture the relation that our true beliefs have to have to reality in order to constitute knowledge. According to the presented view, knowledge is essentially a state in which one could not easily have erred. To know is to be safe from error.

Our main project was to illuminate the nature of knowledge and to find a theory of knowledge that affirms most of our everyday knowledge attributions as true. I think it is safe to say that we have succeeded in these goals at least to some extent. Even if time were to prove our theory wrong, we would still have managed to shed some light on the modal relation that belief has to bear to truth in order to constitute knowledge. To know is to be safe from error.

References:

Alspector-Kelly, M. (2010). "Why Safety Doesn't Save Closure." *Synthese*, 183, 127-142.

Armstrong, D. (1973). *Belief, Truth and Knowledge*. London: Cambridge University Press.

Bauman, P. (2012) "Nozick's Defense of Closure." *Sensitivity Principle in Epistemology.* (Eds.) K. Becker, T. Black, 11-27. New York: Oxford University Press.

Becker, K. (2007). *Epistemology Modalized*. New York: Routledge.

Becker, K. (2012). "Methods and How to Individuate Them". *Sensitivity Principle in Epistemology.* (Eds.) K. Becker, T. Black, 81-97. New York: Oxford University Press.

Black, T. (2002). "A Moorean Response to Brain-in-a –Vat Skepticism." *Australasian Journal of Philosophy*, 80:2, 148-163.

Bogardus, T. (2012). "Knowledge Under Threat." *Philosophy and Phenomenological Research*, 1-25.

Carter, J. A. (2010). “Anti-Luck Epistemology and Safety’s (Recent) Discontents.” *Philosophia*, 38, 517-532.

Chisholm, R. (1977). *Theory of Knowledge.* Englewood Cliffs, NJ: Prentice-Hall.

Coffman, E. J. (2007). “Thinking about luck.” *Synthese,* 158, 385-998.

Comesaña, J. (2005). “Unsafe Knowledge.” *Synthese,* 146, 395-404.

Conee, E. & Feldman, R. (1998). “The Generality Problem for Reliabilism” *Philosophical Studies*, 89, 1-29.

Cohen, S. (2004). “Structure and Connection: Comment’s on Sosa’s Epistemology.” *Ernest Sosa and His Critics*. (Ed.) J. Greco, 17-21. Oxford: Blackwell Publishing Ltd.

Craig, E. (1990). *Knowledge and the State of Nature*. New York: Oxford University Press.

DeRose, K. (1995). “Solving the Skeptical Problem.” *The Philosophical Review*, 104, No. 1, 1-52.

Gettier, E. (1963). “Is Justified True Belief Knowledge?” *Analysis*, 23, 121-123.

Goldman, A. (1976). “Discrimination and Perceptual Knowledge.” *The Journal of Philosophy*, 73, 771-791.

Goldman, A. (1983). “Review of Nozick’s *Philosophical Explanations.*” *Philosophical Review*, 92, No. 1, 81-88.

Greco, J. (2007). “Worries about Pritchard’s Safety.” *Synthese*, 158, 299-302.

Greco, J. (2010). *Achieving Knowledge. A Virtue-Theoretic Account of Epistemic Normativity.* Cambridge: Cambridge University Press.

Greco, J. (2012). “Better Safe than Sensitive.” *The Sensitivity Principle in Epistemology.*(Eds.) Becker, K & Black, T., p. 193-206. New York: Cambridge University Press.

Harman, G. (1973). *Thought*. Princeton: Princeton University Press.

Hawthorne, J. (2004) *Knowledge and Lotteries.* New York: Oxford University Press.

Lackey, J. (2007). “Why We Don’t Deserve Credit for Everything We Know.”. *Synthese*, 158, 345-361.

Lackey, J. (2008). “What luck is not.” *Australasian Journal of Philosophy*, 86, 2, 255-267.

Lehrer, K. & Paxson, T. (1969). “Knowledge: Undefeated Justified True Belief.” *Journal of Philosophy*, 66, 8, 225-237.

Lewis, D. (1973). *Counterfactuals.* Oxford: Blackwell.

Luper-Foy, S. (1984). “The Epistemic Predicament: Knowledge, Nozickian Tracking and Scepticism”. *Australasian Journal of Philosophy*, 62, 26-49.

Luper, S. (2006). “Restorative Rigging and the Safe Indication Account.” *Synthese*, 153, 161-170.

Neta, R. & Rohrbaugh, G. (2004). “Luminosity and the Safety of Knowledge.” *Pacific Philosophical Quarterly*, 85, 396-406.

Nozick, R. (1981). *Philosophical Explanations.* Oxford: Oxford University Press.

Pritchard, D. (2002a). “Recent Work on Radical Skepticism.” *American Philosophical Quarterly*, 39, 215-257.

Pritchard, D. (2002b). “ Resurrecting the Moorean Response to Scepticism.” *International Journal of Philosophical Studies*, 10, 283-307.

Pritchard, D. (2005). *Epistemic Luck.* Oxford: Oxford University Press.

Pritchard, D. (2007). “Anti-luck epistemology.” *Synthese*, 158, 277-297.

Pritchard, D. (2008). “Sensitivity, Safety and Antiluck Epistemology” *The Oxford Handbook of Skepticism*. (Ed.) J. Greco, 1-26, accessed from Oxford Handbooks online 20.3.2013. DOI: 10.1093/oxfordhb/9780195183214.003.0021

Pritchard, D. (2009). *Knowledge*. Hampshire: Palgrave MacMillan.

Pritchard, D. (2010). “Knowledge and Understanding”. *The Nature and Value of Knowledge.* (Eds.) D. Pritchard, A. Millar, A. Haddock, 5-88. Oxford: Oxford University Press.

Pritchard, D. (2012a). “In Defence of Modest Anti-Luck Epistemology”. *Sensitivity Principle in Epistemology.* (Eds.) K. Becker, T. Black, 173-192. New York: Oxford University Press.

Pritchard, D. (2012b). *Epistemological Disjunctivism*. Oxford: Oxford University Press.

Murphy, P. & Black, T. (2012). "Sensitivity Meets Explanation: an Improved Counterfactual Condition on Knowledge." *The Sensitivity Principle in Epistemology.*(Eds.) Becker, K & Black, T., p. 28-42. New York: Cambridge University Press.

Riggs, W. (2007). "Why epistemologists are so down their luck." *Synthese,* 158, 329-344.

Sacks, O. (1985). *The Man Who Mistook His Wife for a Hat and Other Clinical Tales*. London: Duckworth.

Shope, R. (2002). " Conditions and Analyses of Knowledge." *The Oxford Handbook of Epistemology*. (Ed.) P. Moser, 25-71.

Sosa, E. (1999). "How to Defeat Opposition to Moore." *Philosophical Perspectives* 13, 141-153.

Sosa, E. (2004). "Replies." *Ernest Sosa and His Critics*. (Ed.) J. Greco, 275-325. Oxford: Blackwell Publishing Ltd.

Sosa, E. (2009) "Tracking, Competence, and Knowledge." *The Oxford Handbook of Epistemology.* Ed. P. Moser. New York: Oxford University Press. Accessed 18 July 2013. doi: 10.1093/oxfordhb/9780195301700.001.0001

Sosa, E. (2011). *Reflective Knowledge. Apt Belief and Reflective Knowledge, Volume II.* Oxford: Oxford University Press.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

Zalabardo, J. (2012). *Scepticism & Reliable Belief*. Oxford: Oxford University Press.

Appendix of thought experiments: